



OPEN

# Panoramic ultrasound vs. MRI for the assessment of hamstrings cross-sectional area and volume in a large athletic cohort

Martino V. Franchi<sup>1,2✉</sup>, Daniel P. Fitze<sup>2</sup>, Jonas Hanimann<sup>2</sup>, Fabio Sarto<sup>1</sup> & Jörg Spörri<sup>2,3</sup>

We investigated the validity of panoramic ultrasound (US) compared to magnetic resonance imaging (MRI) for the assessment of hamstrings cross-sectional area (CSA) and volume. Hamstrings CSA were acquired with US (by an expert operator) at four different sites of femur length (FL) in 85 youth competitive alpine skiers ( $14.8 \pm 0.5$  years), and successively compared to corresponding scans obtained by MRI, analyzed by a trained vs. a novice rater. The agreement between techniques was assessed by Bland–Altman analyses. Statistical analysis was carried out using Pearson's product moment correlation coefficient ( $r$ ). US-derived CSA showed a very good agreement compared to MRI-based ones. The best sites were 40% FL (0 = mid patellar point) for biceps femoris long head ( $r = 0.9$ ), 50% for semitendinosus ( $r = 0.9$ ), and 30% for semimembranosus ( $r = 0.86$ ) and biceps femoris short head (BFsh,  $r = 0.8$ ). US-based vs. MRI-based hamstrings volume showed an  $r$  of 0.96. Poorer  $r$  values were observed for the novice compared to the trained rater, with the biggest difference observed for BFsh at 50% ( $r = 0.001$  vs.  $r = 0.50$ , respectively) and semimembranosus at 60% ( $r = 0.23$  vs.  $r = 0.42$ , respectively). Panoramic US provides valid CSA values and volume estimations compared to MRI. To ensure optimal US-vs.-MRI agreement, raters should preferably possess previous experience in imaging-based analyses.

The assessment of skeletal muscle size is central within many athletic performance and clinical scenarios. Monitoring changes in muscle size allows to gain more specific insights into the actual status of athletes within their long-term athletic development process or their return-to-sport journeys, when recovering from injuries<sup>1</sup>. The most commonly used measures of muscle size are muscle cross-sectional area (CSA) and muscle volume. Muscle CSA has been widely used in research and sport medicine contexts being strongly related to joint torque production in both quadriceps and hamstrings muscle groups<sup>2,3</sup>. Furthermore, muscle CSA assessment has been recently shown to be useful in hospital settings for predicting survival and risk of treatment failure<sup>4</sup> as well as for improving the prognosis process when evaluated overtime<sup>5</sup>. Similarly, muscle volume is regarded as one of the best predictors of joint torque in humans for both upper and lower limbs of healthy males and sport athletes<sup>6,7</sup>.

The quantification of muscle CSA and volume is usually obtained by imaging techniques, such as computer tomography (CT) and magnetic resonance imaging (MRI). The latter is recognized as the gold standard for clinical and research imaging, providing accurate estimations of muscle size and involving minimal radiation exposure compared to CT scans<sup>8</sup>. However, MRI is not cheap and not always accessible as other imaging techniques<sup>9</sup>, especially in athletic-related settings, when many measures should be acquired over time (and preferably independently of fixed imaging facilities) in order to meticulously monitor the athletic development process or the successfulness of return to sport programs.

In such context, ultrasound (US) could represent a cheaper, yet reliable alternative for the quantification of muscle size. Specifically, after the introduction of panoramic ultrasound imaging (i.e. the Extended Field of View—EFOV technique)<sup>10</sup>, sonography can be adopted to measure larger anatomical structures, such as full

<sup>1</sup>Department of Biomedical Sciences, Institute of Physiology, University of Padua, Padua, Italy. <sup>2</sup>Sports Medical Research Group, Department of Orthopaedics, Balgrist University Hospital, University of Zurich, Zurich, Switzerland. <sup>3</sup>University Centre for Prevention and Sports Medicine, Balgrist University Hospital, University of Zurich, Zurich, Switzerland. ✉email: martino.franchi@unipd.it

muscle CSAs when scans are acquired in the transversal plane<sup>11</sup>, as it has been previously described in two pioneering studies investigating the quadriceps muscle group<sup>12,13</sup>.

In the present investigation we have focused on the use of panoramic US (i.e. compared to MRI-based imaging) for the evaluation of hamstrings muscles CSA and volume in a large athletic cohort. Previously, individual hamstrings muscle CSA and volume have been assessed across maturation and athletic development<sup>14</sup>, while recovering from anterior cruciate ligament reconstruction (ACLR) or hamstring strains<sup>15–17</sup>, and before return-to-sport clearance<sup>1</sup>. Moreover, with respect to both ACLR and hamstring strains, substantial reductions of muscle volume in biceps femoris long head (BF<sub>lh</sub>) or semitendinosus (ST) have been observed, and monitoring muscle size becomes even more imperative in situations where a re-injury may occur<sup>18–20</sup>. In the specific context of competitive alpine skiers, hamstrings muscles play an especially important role, as they may act as an ACL-synergist by producing a posteriorly directed shear force to the tibia (i.e., by eccentrically resisting the boot-induced anterior drawer of the tibia relative to the femur that is known to be typical for skiing-related ACL injury mechanisms)<sup>21–24</sup>. In all of these examples, the US-based evaluation of hamstrings CSA and volume may serve as a meaningful monitoring tool.

Therefore, the main objectives of our study were to investigate the validity of panoramic US compared to MRI for the assessment of (1) individual hamstring muscles CSA and (2) hamstring muscles volume in a large athletic cohort of 85 youth (13–15 years old) competitive alpine skiers. Additional objectives in subgroups of the entire athletic cohort were: (3) to assess the intra-session test–retest reliability of the US-derived CSA acquisition and assessment; (4) to compare the US-derived CSA post-acquisition evaluation (contour tracing) of a trained vs. a complete novice rater (inter-rater reliability); and (5) to explore the differences in US-derived CSA contour tracing of the trained rater before and after the MRI analyses being carried out (effect of image-analysis training).

## Methods

Throughout the manuscript, all values are presented as means ± SD.

We recruited 85 youth competitive alpine ski racers (36 females, 49 males—14.8 ± 0.5 years old, height 166 ± 7.6 cm, body weight 56.6 ± 9 kg) that were part of a certificated regional performance center (RLZ/CRP) of Swiss-Ski, i.e. representing the best level skiers of that age-group in Switzerland. Measurements were acquired during the preseason period (October–December). The protocol was approved by the institutional review board at Balgrist University Hospital and the local ethics committee (KEK-ZH-NR: 2017-01395). All subjects gave written informed consent in accordance with the Declaration of Helsinki.

**Ultrasound imaging.** All ultrasound images were acquired by an expert operator (MVf) with the same ultrasound device throughout the whole study (Aixplorer Ultimate, SuperSonic Imagine, Aix-en-Provence, France) using a linear 50 mm transducer (SuperLinear SL18-5, SuperSonic Imagine, Aix-en-Provence, France). Participants were asked to lie prone on a massage bed and were instructed to rest with extended knee joints, and to relax completely during image acquisition with their feet placed just outside the bed frame in order to avoid any lower limb imbalances. Five minutes of rest in this position were provided for body fluid shift stabilization.

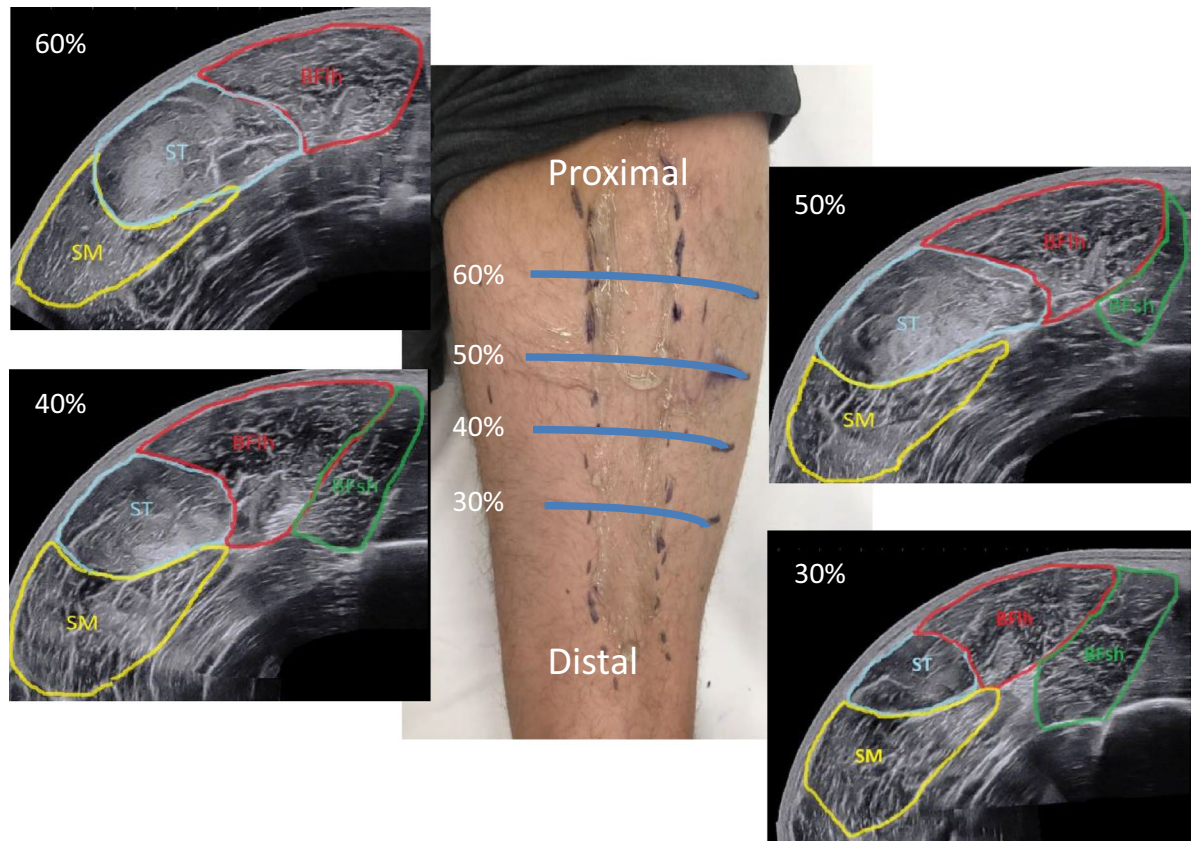
For the panoramic ultrasound scans, the region of interest (ROI) was determined and marked as follows. Firstly, a mark was drawn at 50% of the distance between the greater trochanter and the mid patella point. In second instance, other marks were drawn with a permanent ink pen at the 30%, 40% and 60% of the femur length (distal to proximal, thus 0% was representing the mid patellar point and 100% the great trochanter). The ROIs were identified laterally, from the borders of vastus lateralis muscle, and medially, until the borders of the gracilis muscle, similar to the guidelines suggested by Balias and colleagues<sup>25</sup>. Multiple ROIs were identified in a similar fashion to the study of Kositsky and colleagues<sup>26</sup>, the reason being that distinct muscles present their CSA peak value at distinct muscle sites. In addition, obtaining CSA values at multiple sites allowed us to reconstruct a bigger portion of muscle volume of the hamstrings muscle group.

At each ROI, the transducer was placed on the lateral portion of the posterior thigh (just before the Biceps femoris long and short heads muscle borders) and then moved on the transversal plane in a lateral-to-medial fashion until the end of borders of the semimembranosus and the start of gracilis muscle were identified, then the panoramic acquisition was stopped. We have carefully ensured that the images were collected at the right angle in the transverse plane by using a plastic guide placed on the skin of the volunteer's thigh, similar to the one used by Noorkoiv et colleagues<sup>13</sup>. The transducer was kept in contact with the guide throughout the whole acquisition of CSAs, thus we ensured that the right CSA path was followed while keeping the transducer perpendicular to the skin. The operator took meticulous care in keeping the pressure as constant as possible during the entire image acquisition. For all scans, transmission gel was used to improve the acoustic contact and to keep the pressure on the skin to a minimum. For the purpose of test–retest reliability assessment as outlined below, EFOV scans were acquired twice after completely removing the transducer from the skin and with 5 min rest period in between.

Ultrasound images were analyzed in randomized order by tracing the contours of each four of the hamstring muscles (BF<sub>lh</sub>, BF<sub>sh</sub>, ST, SM) at each ROIs using Image J, a public domain software for image analysis (<https://imagej.nih.gov/ij/>). Once CSAs measurements were obtained at the specific ROIs along the muscle length, by knowing the distance between the different ROIs (constant, as they were identified as percentages of the total femur length), muscle volume (VOL) was calculated for each muscle (for the portion between 30 and 60% of whole femur length) using the truncated cone formula<sup>27,28</sup>:

$$\text{VOL} = 1/3 * h * [\text{CSA}_1 + \text{CSA}_2 + \sqrt{(\text{CSA}_1 * \text{CSA}_2)}]$$

where h is the distance between CSA<sub>1</sub> and CSA<sub>2</sub>.



**Figure 1.** The four measurement sites (ROIs) at different portion of the femur length and the relative panoramic US scans of hamstring muscles. The picture was taken by MVF and DPF at Uniklinik Balgrist, Swiss Centre for Musculoskeletal Imaging, Zurich, Switzerland.

**Magnetic resonance imaging.** All MRI data were acquired on a 3T scanner (Magnetom Prisma, Siemens, Erlangen, Germany). Participants were positioned supine on the MR system's patient bed, fitted with ear protection, and instructed regarding the use of the alarm bell and scans, followed by the acquisition of localizer scans. After ca. 5 min that allowed for body fluid shift stabilization, up to three spatially overlapping T1w spoiled 3D gradient echo data sets of both thighs were acquired in the axial (transverse) orientation and combined into a single image series. Acquisition parameters: sequence: vibe, TR/TE: 3.23/1.23 ms, flip angle: 15°, field of view: 450 × 337.5 mm, encoded nominal voxel size: 1.3 × 1.3 × 3 mm, no inter-slice gap, readout-bandwidth 790 Hz/pixel (278 kHz), acquisition time per 3D volume: 1 min 23 s. In the corresponding images, contours of the semimembranosus (SM), the semitendinosus (ST), and the biceps femoris long (BFh) and short heads (BFsh) muscles were digitized in randomized order using the Merlin Diagnostic Workcenter DICOM image analysis software (Phönix-PACS GmbH, Version 5.3.156494, <https://www.phoenix-pacs.de/image-display/>) (Fig. 1). As for the US technique, once CSAs measurements were obtained at the specific ROIs along the muscle length, muscle volume (VOL) was calculated using the truncated cone formula<sup>27,28</sup> for each muscle (for the portion between 30 and 60% of whole femur length).

**Statistical analyses.** Data analysis was performed using GraphPad Prism software (version 8.00, GraphPad Software, San Diego California USA).

**Comparison of a panoramic US- vs. MRI-derived hamstrings CSA and volume assessment (objective 1 and 2).** US- and MRI-derived hamstrings CSA and hamstrings muscle volume data were reported as mean ± SD. Normality of distribution was checked by the Shapiro–Wilk's test. The agreement between techniques was assessed by using Bland–Altman analyses<sup>29,30</sup>. The validity between MRI and US was tested by plotting the differences in CSA measurements by both techniques against their means, setting ± 1.96 SD as the limits of agreement<sup>12</sup>. If the differences were within ± 1.96 SD, then the two methods can be used with similar accuracy to measure muscle CSA. Correlations were tested by using the Pearson's product moment correlation coefficient ( $r$ ) and their 95% confidence intervals (Table 1). The level of significance was set at  $p < 0.05$ .

**Test–retest reliability US-derived CSA assessment (objective 3).** The intra-session test–retest reliability of hamstrings muscles CSA measured by US was analyzed based on the data of 6 subjects randomly chosen from the entire athletic cohort. The test–retest reliability assessment for EFOV ultrasound technique was performed at

Rater 1—MRI vs. US (MRI CSAs evaluated post US assessment, both in randomized order)					
Muscle (distal—0%, to proximal—100%)	MRI CSA (cm <sup>2</sup> )	Mean bias (cm <sup>2</sup> ) [%]	ULO (cm <sup>2</sup> ) [%]	LLO (cm <sup>2</sup> ) [%]	Pearson's <i>r</i> (95% CI)
<b>Semimembranosus</b>					
30% (n = 85)	10.07 ± 2.32	0.02 ± 1.2 [- 0.4%]	2.38 [24%]	- 2.37 [- 25%]	0.86 (0.79–0.90)
40% (n = 85)	8.09 ± 2.19	- 0.77 ± 1.38 [- 9%]	1.94 [23%]	- 3.5 [- 42%]	0.79 (0.69–0.86)
50% (n = 85)	4.51 ± 1.98	- 1.24 ± 1.37 [- 25%]	1.44 [49%]	- 3.94 [- 99%]	0.78 (0.68–0.85)
60% (n = 70) <sup>a</sup>	1.61 ± 1.15	- 0.99 ± 1.09 [- 53%]	1.15 [52%]	- 3.14 [- 159%]	0.64 (0.48–0.76)
<b>Semitendinosus</b>					
30% (n = 85)	4.17 ± 1.82	0.33 ± 0.96 [3%]	2.23 [74%]	- 1.55 [- 67%]	0.84 (0.77–0.89)
40% (n = 85)	7.33 ± 1.72	0.44 ± 1.04 [7%]	2.48 [38%]	- 1.6 [- 23%]	0.84 (0.77–0.89)
50% (n = 85)	8.64 ± 2	0.26 ± 0.92 [4%]	2.08 [27%]	- 1.55 [- 19%]	0.90 (0.86–0.93)
60% (n = 85)	8.5 ± 2.01	- 0.34 ± 0.99 [- 4%]	1.6 [18%]	- 2.29 [- 26%]	0.87 (0.82–0.92)
<b>Biceps femoris long head</b>					
30% (n = 85)	8.12 ± 2.26	1.19 ± 1.19 [16%]	3.54 [48%]	- 1.15 [- 16%]	0.85 (0.78–0.90)
40% (n = 85)	10.36 ± 2.19	0.77 ± 0.91 [8%]	2.56 [26%]	- 1.02 [- 10%]	0.90 (0.86–0.94)
50% (n = 85)	8.93 ± 1.84	- 0.36 ± 0.98 [- 4%]	1.55 [19%]	- 2.28 [- 27%]	0.83 (0.75–0.89)
60% (n = 84) <sup>a</sup>	5.96 ± 1.73	- 0.91 ± 1.16 [- 15%]	1.35 [22%]	- 3.18 [- 54%]	0.75 (0.64–0.83)
<b>Biceps femoris short head</b>					
30% (n = 85)	5.71 ± 1.38	- 0.08 ± 0.91 [- 1%]	1.7 [34%]	- 1.88 [- 36%]	0.80 (0.71–0.86)
40% (n = 85)	3.71 ± 1.03	- 0.23 ± 0.83 [- 6%]	1.4 [37%]	- 1.88 [- 49%]	0.69 (0.56–0.79)
50% (n = 82) <sup>a</sup>	1.85 ± 0.74	- 0.63 ± 0.83 [- 29%]	1 [40%]	- 2.26 [- 99%]	0.37 (0.17–0.55)
60% (n = 0) <sup>a</sup>	-	-	-	-	-

**Table 1.** Parameters of level of agreement derived from Bland Altman analyses for MRI and US measurements of hamstring muscles CSA at different portions of femur length. Pearson's *r* correlation coefficients are reported too. *ULO* upper limit of agreement, *LLO* lower limit of agreement. <sup>a</sup>Lower number of subjects due to not visible muscle

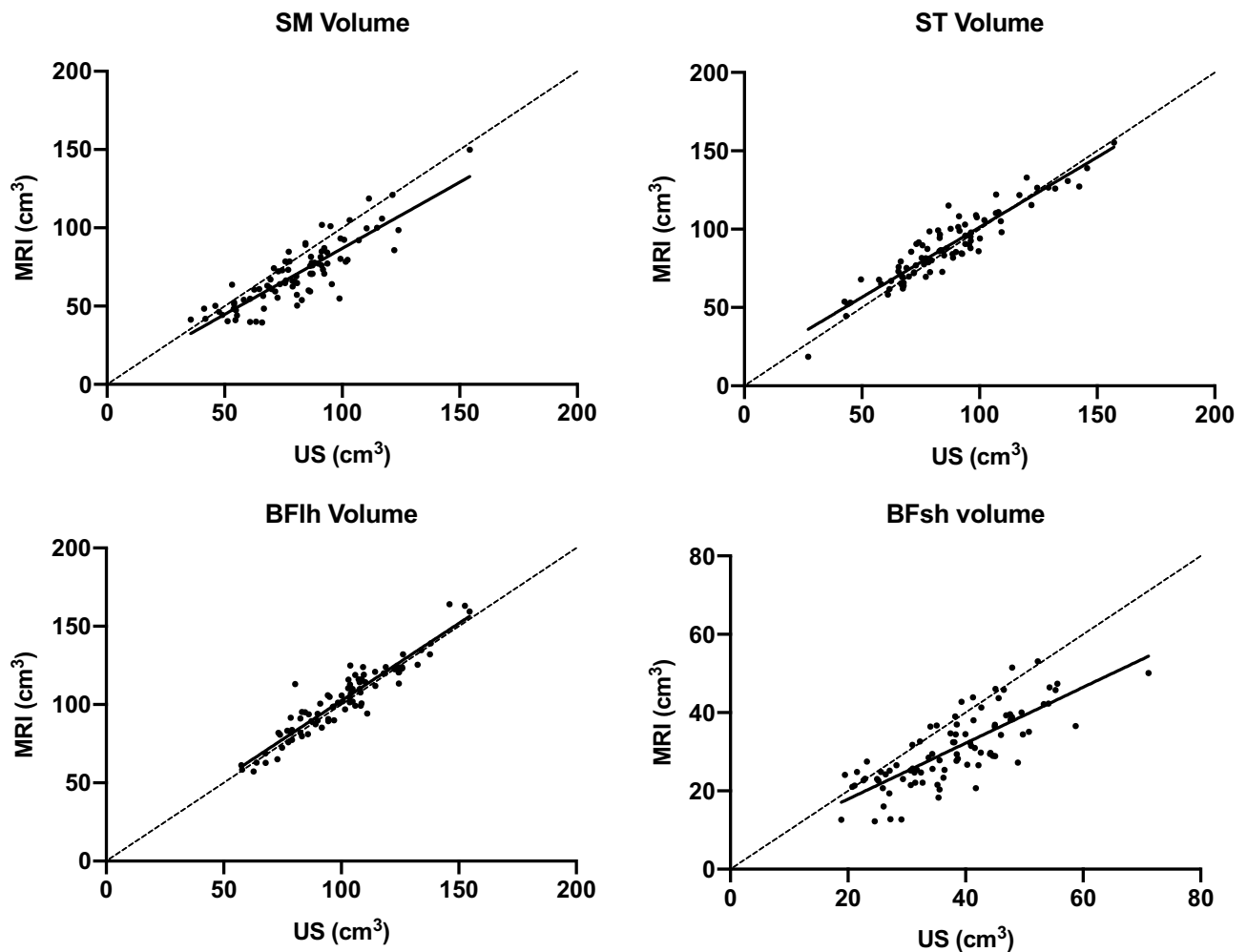
one specific ROI (i.e., the 50% of femur length) by calculating the intraclass correlation coefficient (ICC) and their 95% confidence intervals based on a mean measurement absolute-agreement, two-way mixed model effect (ICC<sub>3,k</sub>), for all muscle architecture parameters, as previously performed by Kositsky and colleagues<sup>26</sup>. Test-retest reliability was classified as good (0.75–0.90) and excellent (>0.90), following the classification proposed by Ko and Li<sup>31</sup>.

*Inter-rater reliability of the US-derived CSA assessment: trained rater vs. complete novel rater (objective 4).* In a subgroup of 45 randomly chosen youth athletes we compared the differences in US-derived CSA contour tracing between two operators (trained—rater 1 vs. completely novel—rater 2) for the same ultrasound images for all the hamstrings muscles at all the ROIs. We decided to select n = 45 youth athletes as the half of the total number (n = 85) of subjects investigated rounded up to n = 45. Image analysis was blinded and performed in a randomized order. The agreement between techniques was assessed by using Bland–Altman analyses for both raters. Correlations were tested by using the Pearson's product moment correlation coefficients (*r*) and their 95% confidence intervals. Further, Pearson's product moment correlation coefficient (*r*) were transformed in Fisher's z-scores and the difference for all z-scores at all muscle sites were tested by paired t-test between the two raters (Table 4). The level of significance was set at *p* < 0.05.

*Differences in US-derived CSA assessment of the trained rater before and after the MRI analyses being carried out (objective 5).* Also, in the same subgroup of 45 randomly chosen youth athletes we compared the differences in US-derived CSA contour tracing of the same rater before and after carrying out the MRI analyses. For clarity, the rater first analyzed US scans, then MRI scans, and lastly US images again. Image analysis was blinded and performed in a randomized order for all the three distinct batches of data analysis. The agreement between techniques was assessed by using Bland–Altman analyses for both rater 1 assessments (pre vs. post MRI analyses). Correlations were tested by using the Pearson's product moment correlation coefficient (*r*) and their 95% confidence intervals. Further, Pearson's product moment correlation coefficients (*r*) were transformed in Fisher's z-scores and the difference for all z-scores at all muscle sites were tested by paired t-test between the two raters (Table 4). The level of significance was set at *p* < 0.05.

## Results

**Comparison of a panoramic US- vs. MRI-derived hamstrings CSA and volume assessment.** The average CSA values for each muscle at each specific ROI (evaluated by a trained rater—rater 1) are presented for MRI and US in Table 1 together with the results of the Bland Altman (mean bias and limits of agreements) and Pearson's correlation analyses. In general, a very good agreement was observed between the two techniques



**Figure 2.** Correlations between MRI and US measurements for the assessment of muscle volume of each hamstring muscles. The volume is representative of the muscle portions scanned, thus between 30 and 60% of femur length.

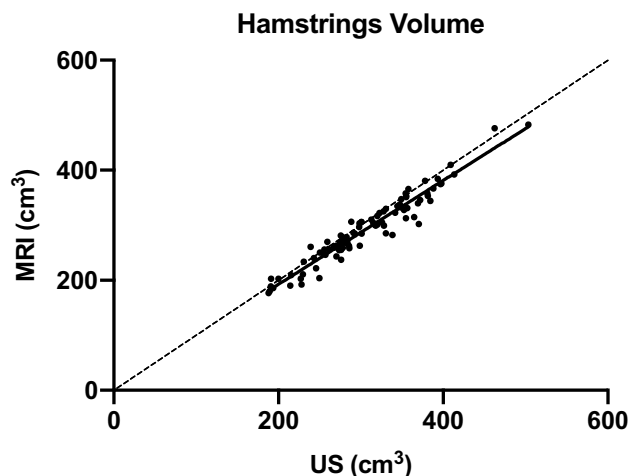
for each hamstrings muscle. Pearson's correlation coefficient ranged from 0.64–0.86 for SM ( $p < 0.0001$  for all ROIs), from 0.84–0.90 for ST ( $p < 0.0001$  for all ROIs), from 0.75–0.90 for BFHh ( $p < 0.0001$  for all ROIs), and from 0.37–0.80 for BFsh ( $p < 0.0001$  for 30% and 40% ROIs,  $p < 0.01$  for 50% ROI).

ST and BFHh were the muscles that presented the highest  $r$  values at the ROIs of 50% and 40% of femur length, respectively ( $r = 0.90$  for both muscles at those specific ROIs). The muscles that presented the lowest  $r$  values between US and MRI were SM and BFsh at the ROIs of 60% and 50% of the femur length, respectively ( $r = 0.64$  and  $r = 0.37$ , mean % bias =  $-53\%$  and  $-29\%$ , respectively). Over the specific ROI in which the scans were acquired (between 30 and 60% of femur length, mid patellar point regarded as 0%), maximal CSA measured from MRI was  $8.09 \pm 2.19$  for SM located at the 40%,  $8.64 \pm 2$  for ST located at the 50%,  $10.36 \pm 2.19$  for BFHh located at the 50%, and  $5.71 \pm 1.38$  for BFsh located at the 30%.

The comparisons between volume calculations (for the muscle portions scanned between 30 and 60% of femur length) carried out by rater 1 from all the US vs. MRI-derived CSA analyses are presented for each muscle (Fig. 2) and for all the hamstrings pooled together (Fig. 3). Pearson's  $r$  values were  $r = 0.86$  for SM,  $r = 0.93$  for ST,  $r = 0.93$  for BFHh, and  $r = 0.78$  for BFsh. For all the hamstrings pooled together, the  $r$  value for hamstrings volume was 0.96 between US and MRI analyses.

**Test–retest reliability US-derived CSA assessment.** The intra-session reliability data ( $ICC_{3,k}$ , Pearson's correlations and SEMs) for each muscle CSA acquired by EFOV US technique at the 50% of femur length are presented in Table 2. SM, ST and BFHh muscles showed a very high relative reliability ( $ICC_{3,k}$  95% CI 0.74–0.99, 0.93–0.99, and 0.87–0.99, respectively). A high repeatability was also observed for BFsh CSA measurement ( $ICC_{3,k}$  95% CI 0.36–0.98).

**Inter-rater reliability of the US-derived CSA assessment.** The US-derived CSA values and the Pearson's correlation coefficients for rater 1 (trained) and rater 2 (completely novel) (relative US vs. MRI comparison)



**Figure 3.** Correlation between MRI and US measurements for the assessment of muscle volume of hamstring muscles pooled together. The volume is representative of the muscle portions scanned, thus between 30 and 60% of femur length.

Muscle	Measure1 (cm <sup>2</sup> )	Measure2 (cm <sup>2</sup> )	ICC <sub>3,k</sub> (95% CI)	SEM (cm <sup>2</sup> )
SM	6.11 ± 1.57	5.94 ± 1.46	0.96 (0.74–0.99)	0.41
ST	9.53 ± 2.88	9.77 ± 2.95	0.99 (0.93–0.99)	0.19
BFlh	9.58 ± 2.25	9.64 ± 1.94	0.97 (0.87–0.99)	0.57
BFsh	2.21 ± 0.61	1.97 ± 0.6	0.85 (0.36–0.98)	0.17

**Table 2.** Intra-session repeatability for CSA values for each muscle at the 50% of femur length region of interest, measured by EFOV technique within the same session (n = 6, 15 min between each measure). ICC intraclass correlation coefficient (with 95% confidence intervals), SEM standard error of the mean.

are shown in Table 3 for a sub-cohort of 45 athletes. Generally, worse Pearson's *r* values were observed for rater 2 compared to rater 1, the biggest difference being represented by BFsh values at 30% of femur length ( $r = 0.001$  vs.  $r = 0.50$ , respectively) and SM at 60% ( $r = 0.23$  vs.  $r = 0.42$ , respectively). The Fisher's *z*-score for the same analyses are shown in Table 4; rater 1 shows a higher overall *z*-score compared to rater 2 ( $p < 0.001$ ).

#### Differences in US-derived CSA assessment of a trained rater before and after the MRI analyses being carried out.

The mean biases (with lower and upper limits of agreement) and the Pearson's correlation coefficients for the trained rater's second US-derived CSA assessment and MRI (post MRI analyses) are also shown in Table 3, being carried out for each muscles and ROIs. Generally, after MRI evaluation better Pearson's *r* values were observed, the biggest improvement being observed on BFsh at 30% and 50% ROIs ( $r = 0.79$  vs.  $r = 0.91$ , and  $r = 0.50$  vs.  $r = 0.82$ , respectively). The Fisher's *z*-score for the same analyses are shown in Table 4; assessment number 2 (post MRI analyses) shows a higher overall *z*-score compared to assessment 1 ( $p < 0.001$ ).

#### Discussion

The main findings of the present study pointed out that US is valid and reliable for the assessment of individual hamstrings CSA and volume when compared to the gold-standard MRI. The Bland Altman analyses showed that the trained rater presented a better agreement (i.e., lower biases and narrower limits of agreements) between MRI and US-derived CSA values compared to a completely novice rater; further, it appears that is preferable to have previous experience in MRI-derived CSA evaluation.

#### Methodological considerations: a valid and reliable assessment of US-derived hamstring CSA and volume assessment appears feasible.

In first instance, our results revealed a good agreement between a panoramic US- vs. MRI-derived hamstrings CSA and volume assessments. Thus, compared to the golden reference standard MRI, a US-based approach can be considered being sufficiently valid. Moreover, the CSA agreement was better for specific muscles (ST and BFlh) at specific sites (50% and 40% of femur length, respectively). These data strongly support the previously published observations of Kositsky and colleagues<sup>26</sup>. In a very similar fashion, they showed that the highest agreement between the two imaging techniques was found at ROIs where CSA were the largest: in fact, in Kositsky et al. study the poorest agreement was found for BFsh muscle, especially closer to the 50% ROI ( $r = -0.09$ ), which was similar in the present investigation ( $r = 0.37$ ). As

Rater 1 (trained) vs. Rater 2 (novice) (US) (MRI CSAs assessed in randomized order by rater 1 post US assessment n1)					Rater 1 (trained) assessment n2 (US) (US CSAs re-assessed in randomized order post MRI assessment)	
Muscle (distal—0%, to proximal—100%)	Rater 1 Mean bias (LLOA – ULOA) (cm <sup>2</sup> )	Rater 1 Pearson's r (95% CI) (US vs. MRI)	Rater 2 Mean bias (LLOA – ULOA) (cm <sup>2</sup> )	Rater 2 Pearson's r (95% CI) (US vs. MRI)	Rater 1 US CSA—n2 Mean bias (LLOA – ULOA) (cm <sup>2</sup> )	Rater 1 Pearson's r (95% CI) (US-2 vs. MRI)
<b>Semimembranosus</b>						
30% (n = 45)	-0.17 (-2.54 to 2.20)	0.90 (0.83 to 0.94)	-0.59 (-4.36 to 3.18)	0.76 (0.59 to 0.86)	0.09 (-1.54 to 1.73)	0.95 (0.92 to 0.97)
40% (n = 45)	-0.53 (-3.02 to 1.94)	0.85 (0.75 to 0.92)	-0.74 (-3.51 to 2.01)	0.85 (0.73 to 0.91)	-0.53 (-1.80 to 1.23)	0.94 (0.90 to 0.97)
50% (n = 45)	-1.01 (-3.70 to 1.78)	0.81 (0.69 to 0.89)	-0.96 (-4.97 to 3.03)	0.69 (0.49 to 0.81)	-0.59 (-2.66 to 1.58)	0.86 (0.75 to 0.92)
60% (n = 40) <sup>a</sup>	-1.04 (-3.51 to 1.41)	0.42 (0.13 to 0.65)	-1.16 (-4.36 to 2.02)	0.23 (-0.01 to 0.50)	-1.12 (-3.71 to 1.46)	0.41 (0.11 to 0.64)
<b>Semitendinosus</b>						
30% (n = 45)	0.19 (-1.73 to 2.12)	0.84 (0.73 to 0.91)	0.68 (-1.37 to 2.75)	0.83 (0.70 to 0.90)	0.38 (-1.06 to 1.83)	0.91 (0.85 to 0.95)
40% (n = 45)	0.39 (-1.90 to 2.69)	0.84 (0.72 to 0.91)	0.61 (-1.52 to 2.74)	0.87 (0.78 to 0.93)	0.51 (-1.25 to 2.29)	0.90 (0.83 to 0.94)
50% (n = 45)	0.24 (-1.54 to 2.06)	0.92 (0.87 to 0.95)	0.55 (-1.87 to 2.98)	0.88 (0.79 to 0.93)	0.29 (-1.01 to 1.58)	0.95 (0.92 to 0.97)
60% (n = 45)	-0.19 (-2.07 to 1.67)	0.90 (0.83 to 0.94)	-0.22 (-3.06 to 2.61)	0.77 (0.62 to 0.87)	-0.23 (-1.84 to 1.36)	0.93 (0.87 to 0.96)
<b>Biceps femoris long head</b>						
30% (n = 45)	1.21 (-0.73 to 3.16)	0.89 (0.81 to 0.94)	0.84 (-1.92 to 3.60)	0.82 (0.68 to 0.89)	1.09 (-0.94 to 3.14)	0.88 (0.80 to 0.93)
40% (n = 45)	0.75 (-1.42 to 2.93)	0.89 (0.81 to 0.94)	0.24 (-2.12 to 2.60)	0.87 (0.77 to 0.92)	0.89 (-0.87 to 2.66)	0.92 (0.87 to 0.96)
50% (n = 45)	-0.22 (-2.19 to 1.74)	0.88 (0.78 to 0.93)	-1.11 (-3.54 to 1.31)	0.84 (0.72 to 0.91)	-0.07 (-1.50 to 1.35)	0.93 (0.88 to 0.96)
60% (n = 45)	-0.69 (-3.19 to 1.79)	0.76 (0.60 to 0.86)	-1.77 (-4.68 to 1.13)	0.72 (0.53 to 0.83)	-0.79 (-2.97 to 1.38)	0.82 (0.70 to 0.90)
<b>Biceps femoris short head</b>						
30% (n = 45)	0.02 (-1.89 to 1.93)	0.79 (0.65 to 0.88)	-0.15 (-2.32 to 2.01)	0.78 (0.63 to 0.87)	-0.27 (-1.42 to 0.86)	0.91 (0.85 to 0.95)
40% (n = 45)	-0.08 (-1.51 to 1.33)	0.74 (0.57 to 0.84)	-0.04 (-1.87 to 1.78)	0.64 (0.42 to 0.78)	-0.02 (-1.44 to 1.39)	0.78 (0.64 to 0.87)
50% (n = 36) <sup>a</sup>	-0.65 (-1.98 to 0.67)	0.50 (0.21 to 0.71)	-0.68 (-3.1 to 1.73)	-0.001 (-0.03 to 0.32)	-0.25 (-1.28 to 0.77)	0.82 (0.67 to 0.90)
60% (n = 0) <sup>a</sup>		-	-	-	-	-

**Table 3.** (Left portion) Comparison of the digitation of muscle contours from US between a trained and a novice rater. Presented are the absolute levels of agreement derived from Bland Altman analyses (with lower and upper limits of agreement) and Pearson's correlation coefficients (r) (with 95% confidence intervals) of MRI and US measurements for hamstring muscles CSA at different portions of femur length. (Right portion) Absolute level of agreement and Pearson's correlation coefficients (r) (with 95% confidence intervals) between MRI and US measurements for the trained rater after the analysis of MRI scans. <sup>a</sup>Lower number of subjects due to not visible muscle.

Kositsky and colleagues explained in the discussion section of their study<sup>26</sup>, MRI shows low minimal detectable change (< ~7%) when assessing small muscles<sup>32</sup>. Therefore, errors in MRI-based CSA evaluation are unlikely to justify the large limits of agreements observed both in their and our studies. Nevertheless, the large relative differences in muscle CSA would be only found at sites where the CSA would be small, and this would represent less of a problem at ROIs of large CSA: in fact, better agreement was found for each muscle in the present study at large CSA sites (e.g. SM and BFsh at 30% of the femur length). In contrast with the previous work of Kositsky and colleagues<sup>26</sup>, we often observed a higher bias towards larger US values compared to MRI ones, especially for muscles that were more difficult to identify with US, usually for the presence of small image artifacts at specific ROIs, such as SM (at 50% and 60% and BFsh at 50%), which likely prevented the accurate identification and digitization of the CSA path. Therefore, we agree with Kositsky et al. that “interchanging and direct comparison of US and MRI measures from literature is inadvisable”.

With respect to reliability, the intra-session test–retest reliability US-derived CSA assessment was shown to be good to excellent (ICC ranging from 0.85 to 0.99) and, thus, confirmed previously reported ICCs for all hamstring muscles (0.88–0.99)<sup>26</sup> and for ST, SM, and BFsh (0.92–0.98)<sup>33</sup>. Furthermore, in a sub-cohort of 45 athletes, the US vs. MRI agreement was observed to be higher for the trained vs. the novice rater, but, remarkably, the agreement also improved for the trained rater after carrying out the MRI analyses. Noteworthy, from the Bland Altman analyses we can notice that the difference of the mean biases between rater 1 and 2 was not observed to be as large as the ranges of the limits of agreement, suggesting that rater 2 was more prone to produce inconsistent US CSA traces when compared to MRI. This is supported by the significant difference found in the Fisher's z-score for all the analyzed CSAs between rater 1 and 2 (Table 4). This methodological aspect acquires great practical impact, as novel operators should refrain from carrying out studies or clinical assessments if not properly trained. Moreover, it appears that previous experience with MRI-based analyses could represent an advantage even for trained ultrasonography operators.

**Sports practical and clinical importance: adding a seat to the table for panoramic US-derived hamstring CSA and volume assessments.** The assessment of individual hamstring muscle CSA by MRI has been used in clinical and sportive settings in order to monitor either the successfulness of post-surgery

Fisher's z-scores			
Rater 1, rater 2, rater 1 assessment n2 (post MRI assessment)			
Muscle (distal—0%, to proximal—100%)	Rater 1 (trained)	Rater 2 (novice)	Rater 1 assessment n2
<b>Semimembranosus</b>			
30% (n = 85)	1.47	0.97	1.83
40% (n = 85)	1.25	1.22	1.73
50% (n = 85)	1.12	0.84	1.29
60% (n = 70) <sup>a</sup>	0.44	0.23	0.43
Mean ± SD	1.07 ± 0.44	0.81 ± 0.41	1.32 ± 0.63
<b>Semitendinosus</b>			
30% (n = 85)	1.22	1.18	1.52
40% (n = 85)	1.22	1.33	1.47
50% (n = 85)	1.58	1.37	1.83
60% (n = 85)	1.47	1.02	1.66
Mean ± SD	1.37 ± 0.18	1.22 ± 0.16	1.62 ± 0.16
<b>Biceps femoris long head</b>			
30% (n = 85)	1.42	1.12	1.37
40% (n = 85)	1.42	1.33	1.59
50% (n = 85)	1.37	1.22	1.65
60% (n = 84) <sup>a</sup>	0.99	0.88	1.15
Mean ± SD	1.30 ± 0.20	1.14 ± 0.19	1.44 ± 0.22
<b>Biceps femoris short head</b>			
30% (n = 85)	1.07	1.04	1.52
40% (n = 85)	0.95	0.75	1.04
50% (n = 82) <sup>a</sup>	0.54	0.01	1.15
Mean ± SD	0.85 ± 0.27	0.60 ± 0.53	1.24 ± 0.25
All muscles	1.17 ± 0.33***	0.97 ± 0.39	1.42 ± 0.36***,###
Mean ± SD			

**Table 4.** Fisher's z-scores (obtained from the transformation of the Pearson's r values presented in Table 3) for the trained rater (rater 1), the novice rater (rater 2), and for the trained rater after MRI assessment. \*\*\* p < 0.001 vs. rater 2, ### p < 0.001 for rater 1 n2 (post-MRI) vs. rater 1. <sup>a</sup>Lower number of subjects due to not visible muscle.

treatments (e.g., after anterior cruciate ligament reconstruction involving a hamstring muscle graft<sup>15,19,20,34</sup>, or to identify potential risks of injury<sup>17</sup>. In such contexts, a US-derived hamstring CSA and volume assessment, possibly carried out with portable devices (and thus independent of imaging facilities), may open new athlete monitoring perspectives and possibilities. Specifically, for the cohort of the present study, physical aspects of the athlete have been suggested to be among the top 5 key injury risk factors in alpine ski racing<sup>35</sup>. During the “landing back weighted” ACL-rupture mechanisms, internal developed forces from the quadriceps muscles may strain the ACL in the distal range of motion close to full knee extension<sup>36,37</sup>, while the hamstring muscles act as an ACL synergist by producing a posteriorly directed shear moment on the tibia<sup>38,39</sup>. Thus, accurate values of muscle CSA can be very useful, being muscle strength previously related to changes in muscle size<sup>2,3</sup>. In fact, Konishi and Fukubayashi showed that the muscle torque per unit volume of patients with ACLR at 12 months in both injured, and uninjured sides was significantly lower than those of controls<sup>40</sup>. Furthermore, a marked reduction of ST volume (~ - 87%) was found in the operated limb and was significantly smaller than those in the contralateral limb after ACLR<sup>20</sup> and this could remain altered from 1 up to 6 years from ACLR<sup>19</sup>.

Although maximal CSA of individual hamstring muscle appears to be an important index for injury prevention and/or post-surgery monitoring, the calculation of total or partial muscle volume could be regarded as an even more significant measure, as it considers some of the regionality of hypertrophic or atrophic processes<sup>9</sup>. In Fig. 3 it is shown that the individual muscle volume calculated for the portion belonging to the 30–60% of the femur length can be reliably obtained with US when compared to MRI (BFsh showing the least agreement with MRI-based calculations).

**Limitations.** The present study has some limitations. When acquiring scans with panoramic US, the results may be influenced by the use of linear transducers on curved surfaces and by the pressure exerted by the operator on the skin and muscle tissue<sup>11</sup>. Although we took care of keeping the transducer always in contact with the skin and keeping the same pressure throughout the whole scan, image distortion may still have occurred. Another limitation is represented by the fact that we assessed only limited muscle portions belonging to the 30–60% of the femur length. However, at such muscle sites the largest CSAs of each individual hamstring muscle are generally located<sup>25,26</sup>. One further limitation could be represented by the choice of focusing on 13–15-year-old athletes. However, we could argue that youth age is clinically wise the most interesting (i.e., onset of over-use injury developments on tendons, first peak in injury rates). Moreover, due to huge difference in biological



maturation around the growth spurt, variability among subjects might be bigger than later during the elite stage, thus representing a good sample to investigate the validity and accuracy of US vs. MRI for hamstrings CSA assessment. Nevertheless, as a future direction, further studies should be focused on cohorts of different age and belonging to different sports.

In conclusion, panoramic US is a valid and reliable method for the assessment of individual hamstring muscles CSA at different scan sites along the muscle length. Moreover, it can also provide valid estimations of hamstring muscles volume compared to MRI. A higher agreement (with narrower limits of agreement range) between US and MRI was observed for a trained compared to a novel rater, and, in addition, this agreement was found to improve for the trained rater after the analysis of MRI scans. Our results suggest that panoramic US can be applied in sport and clinical scenarios for the assessment of hamstring muscles size, nonetheless a meticulous training for raters is warranted in order to provide measurements which are truly reliable.

Received: 2 May 2020; Accepted: 4 August 2020

Published online: 24 August 2020

## References

- Sanfilippo, J. L., Slider, A., Sherry, M. A., Tuite, M. & Heiderscheid, B. C. Hamstring strength and morphology progression after return to sport from injury. *Med. Sci. Sports Exerc.* **45**, 448–454 (2014).
- Blazevich, A. J., Coleman, D. R., Horne, S. & Cannavan, D. Anatomical predictors of maximum isometric and concentric knee extensor moment. *Eur. J. Appl. Physiol.* **105**, 869–878 (2009).
- Evangelidis, P. E., Massey, G. J., Pain, M. T. G. & Folland, J. P. Strength and size relationships of the quadriceps and hamstrings with special reference to reciprocal muscle balance. *Eur. J. Appl. Physiol.* **116**, 593–600 (2016).
- Scott, J. M. *et al.* Panoramic ultrasound: a novel and valid tool for monitoring change in muscle mass. *J. Cachexia Sarcopenia Muscle* **8**, 475–481 (2017).
- Kuroki, L. M. *et al.* Pre-operative assessment of muscle mass to predict surgical complications and prognosis in patients with endometrial cancer. *Ann. Surg. Oncol.* **22**, 972–979 (2015).
- Fukunaga, T. *et al.* Muscle volume is a major determinant of joint torque in humans. *Acta Physiol. Scand.* **172**, 249–255 (2001).
- Maden-Wilkinson, T. M., Balshaw, T. G., Massey, G. & Folland, J. P. What makes long-term resistance-trained individuals so strong? A comparison of skeletal muscle morphology, architecture, and joint mechanics. *J. Appl. Physiol.* <https://doi.org/10.1152/jappphysiol.00224.2019> (2019).
- Heymsfield, S. B., Adamek, M., Gonzalez, M. C., Jia, G. & Thomas, D. M. Assessing skeletal muscle mass: Historical overview and state of the art. *J. Cachexia Sarcopenia Muscle* **5**, 9–18 (2014).
- Franchi, M. V. *et al.* Muscle thickness correlates to muscle cross-sectional area in the assessment of strength training-induced hypertrophy. *Scand. J. Med. Sci. Sport.* **28**, 846–853 (2018).
- Weng, L. *et al.* US extended-field-of-view imaging technology. *Radiology* **203**, 877–880 (1997).
- Franchi, M. V. *et al.* Muscle architecture assessment: strengths, shortcomings and new frontiers of in vivo imaging techniques. *Ultrasound Med. Biol.* **44**, 2492–2504 (2018).
- Ahtiainen, J. P. *et al.* Panoramic ultrasonography is a valid method to measure changes in skeletal muscle cross-sectional area. *Eur. J. Appl. Physiol.* **108**, 273–279 (2010).
- Noorkoiv, M., Nosaka, K. & Blazevich, A. J. Assessment of quadriceps muscle cross-sectional area by ultrasound extended-field-of-view imaging. *Eur. J. Appl. Physiol.* **109**, 631–639 (2010).
- Davidson, S. P. & McLean, S. G. Effects of maturation on combined female muscle strength and ACL structural factors. *J. Sci. Med. Sports* **19**, 553–558 (2016).
- Konrath, J. M. *et al.* Morphologic characteristics and strength of the hamstring muscles remain altered at 2 years after use of a hamstring tendon graft in anterior cruciate ligament reconstruction. *Am. J. Sports Med.* **44**, 2589–2598 (2016).
- Setuain, I. *et al.* Differential effects of two rehabilitation programs following anterior cruciate ligament reconstruction. *J. Sports Rehabil.* **26**, 544–555 (2016).
- Bourne, M. N., Opar, D. A., Williams, M. D., Najjar, A. & Shield, A. J. Muscle activation patterns in the Nordic hamstring exercise: impact of prior strain injury. *Scand. J. Med. Sci. Sports* **26**, 666–674 (2016).
- Silder, A., Heiderscheid, B. C., Thelen, D. G., Enright, T. & Tuite, M. J. MR observations of long-term musculotendon remodeling following a hamstring strain injury. *Skelet. Radiol.* **37**, 1101–1109 (2008).
- Messer, D. J., Shield, A. J., Williams, M. D., Timmins, R. G. & Bourne, M. N. Hamstring muscle activation and morphology are significantly altered 1–6 years after anterior cruciate ligament reconstruction with semitendinosus graft. *Knee Surg. Sports Traumatol. Arthrosc.* **28**, 733–741 (2019).
- Nomura, Y., Kuramochi, R. & Fukubayashi, T. Evaluation of hamstring muscle strength and morphology after anterior cruciate ligament reconstruction. *Scand. J. Med. Sci. Sports* **25**, 301–307 (2015).
- Bere, T. *et al.* A systematic video analysis of 69 injury cases in World Cup alpine skiing. *Scand. J. Med. Sci. Sports* **24**, 667–677 (2014).
- Bere, T. *et al.* Mechanisms of anterior cruciate ligament injury in world cup alpine skiing: a systematic video analysis of 20 cases. *Am. J. Physiol. Physiol.* **39**, 1421–1429 (2011).
- Semadeni, R. & Schmitt, K. U. Numerical simulations to assess different rehabilitation strategies after ACL rupture in a skier. *J. Sports Rehabil.* **18**, 427–437 (2009).
- Heinrich, D., Van Den Bogert, A. J. & Nachbauer, W. Peak ACL force during jump landing in downhill skiing is less sensitive to landing height than landing position. *Br. J. Sports Med.* **52**, 1086–1090 (2018).
- Balius, R., Pedret, C., Iriarte, I., Sáiz, R. & Cerezal, L. Sonographic landmarks in hamstring muscles. *Skelet. Radiol.* **48**, 1675–1683 (2019).
- Kositsky, A. *et al.* Reliability and validity of ultrasonography for measurement of hamstring muscle and tendon cross-sectional area. *Ultrasound Med. Biol.* **46**, 55–63 (2020).
- Nordez, A. *et al.* Comparison of methods to assess quadriceps muscle volume using magnetic resonance imaging. *J. Magn. Reson. Imaging* **30**, 1116–1123 (2009).
- Tracy, B. L. *et al.* A more efficient magnetic resonance imaging-based strategy for measuring quadriceps muscle volume. *Med. Sci. Sports Exerc.* **35**, 425–433 (2003).
- Giavarina, D. Understanding Bland Altman analysis. *Biochem. Med.* **25**, 141–151 (2015).
- Bland, J. & Altman, D. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **8**, 307–310 (1986).
- Koo, T. & Li, M. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* **15**, 155–163 (2016).

32. Smeulders, M. J. C. *et al.* Reliability of in vivo determination of forearm muscle volume using 3.0 T magnetic resonance imaging. *J. Magn. Reson. Imaging* **31**, 1252–1255 (2010).
33. Palmer, T. B., Akehi, K., Thiele, R. M., Smith, D. B. & Thompson, B. J. Reliability of panoramic ultrasound imaging in simultaneously examining muscle size and quality of the hamstring muscles in young, healthy males and females. *Ultrasound Med. Biol.* **41**, 675–684 (2015).
34. Norte, G. E. *et al.* MRI-based assessment of lower-extremity muscle volumes in patients before and after ACL reconstruction. *J. Sports Rehabil.* **27**, 201–212 (2018).
35. Spörri, J., Kröll, J., Amesberger, G., Blake, O. M. & Müller, E. Perceived key injury risk factors in World Cup alpine ski racing—an explorative qualitative study with expert stakeholders. *Br. J. Sports Med.* **46**, 1059–1064 (2012).
36. Herzog, W. & Read, L. J. Lines of action and moment arms of the major force-carrying structures crossing the human knee joint. *J. Anat.* **182**(2), 213–230 (1993).
37. DeMorat, G., Weinhold, P., Blackburn, T., Chudik, S. & Garrett, W. Aggressive quadriceps loading can induce noncontact anterior cruciate ligament injury. *Am. J. Sports Med.* **32**, 477–483 (2004).
38. Baratta, R. *et al.* Muscular coactivation. The role of the antagonist musculature in maintaining knee stability. *Am. J. Sports Med.* **16**, 113–122 (1988).
39. Solomonow, M. *et al.* The synergistic action of the anterior cruciate ligament and thigh muscles in maintaining joint stability. *Am. J. Sports Med.* **15**, 207–213 (1987).
40. Konishi, Y. *et al.* Relationship between quadriceps femoris muscle volume and muscle torque after anterior cruciate ligament rupture. *Knee Surg. Sports. Traumatol. Arthrosc.* **19**, 641–645 (2011).

## Acknowledgements

The authors would like to thank Swiss Ski and Balgrist foundation stiftung for the financial support for the present study.

## Author contributions

M.V.F. and J.S. conceived the study design idea and the manuscript; M.V.F. and D.P.F. conducted the data acquisition; M.V.F., D.P.F., J.H. and F.S. performed the data analysis; M.V.F., F.S. and J.S. drafted the manuscript; all authors approved the final version of the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to M.V.F.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020