



OPEN

## Test–retest reliability of splenic volume assessment by ultrasonography

Pontus Holmström<sup>1✉</sup>, Frank Pernet<sup>1</sup> & Erika Schagatay<sup>1,2</sup>

While MRI and CT are the gold standards for assessments of splenic size in clinical settings, ultrasonography is particularly suited due to its portability, cost efficiency and easy utilization. However, ultrasonography is associated with subjective assessment, potentially resulting in increased variation. We used a test–retest design aiming to determine the reliability of splenic measurements assessed by ultrasonography during apnea. In addition, we compared reliability between different equations for volume calculations: Koga, Prolate ellipsoid and Pilström. Twelve healthy participants (6 women) performed two tests separated by 15 min, comprising a maximal voluntary apnea in a seated position. Splenic dimensions were measured via ultrasonography for 5 min before and immediately following apnea. Resting splenic volume displayed high test–retest reliability between tests (Pilström:  $157 \pm 39$  mL vs  $156 \pm 34$  mL,  $p = .651$ , ICC = .970,  $p < .001$ , CV =  $2.98 \pm 0.1\%$ ; Prolate ellipsoid:  $154 \pm 37$  mL vs  $144 \pm 43$  mL,  $p = .122$ , ICC = .942,  $p < .001$ , CV =  $5.47 \pm 0.3\%$ ; Koga:  $142 \pm 37$  mL vs  $140 \pm 59$  mL,  $p = .845$ , ICC = .859,  $p < .001$ , CV =  $9.72 \pm 1.4\%$ ). Apnea-induced volumes displayed similar reliability ( $127 \pm 29$  mL vs  $129 \pm 28$  mL,  $p = .359$ , ICC = .967,  $p < .001$ , CV =  $3.14 \pm 3.1\%$ ). Reliability was also high between equations (Pilström vs Prolate ellipsoid: ICC = .818,  $p < .001$ , CV =  $7.33 \pm 0.3\%$ , bias =  $-3.1$  mL, LoA =  $-46.9$  to  $40.7$  mL; Pilström vs Koga: ICC = .618,  $p < .01$ , CV =  $11.83 \pm 1.1\%$ , bias =  $-14.8$  mL, LoA =  $-76.9$  to  $47.3$  mL). We conclude that splenic ultrasonographic measurements have practical applications during laboratory and field-based research as a reliable method detecting splenic volume change consistently between repeated tests. The Pilström equation displayed similar reliability compared to the prolate ellipsoid formula and slightly higher compared to the Koga formula and may be particularly useful to account for individual differences in splenic dimensions.

### Abbreviations

MRI	Magnetic resonance imaging
CT	Computed tomography
ICC	Intraclass correlation coefficient
CV	Coefficient of variation
LoA	Limits of agreement
RBC	Red blood cells
yrs	Years
cm	Centimeters
kg	Kilograms
m	Meter
SpO <sub>2</sub>	Peripheral oxygen saturation
L	Splenic length
T	Splenic thickness
W	Splenic width
$V_{spleen}$	Spleen volume
SD	Standard deviation
CI	Confidence interval
TE	Typical error
mL	Milliliter

<sup>1</sup>Environmental Physiology Group, Department of Health Sciences, Mid Sweden University, Kungälvsgatan 4, 83140 Östersund, Sweden. <sup>2</sup>Swedish Winter Sports Research Centre, Department of Health Sciences, Mid Sweden University, Östersund, Sweden. ✉email: Pontus.holmstrom@miun.se

The spleen is a highly vascularized organ with a continuous blood flow pulsating through it<sup>1</sup>, which is important to initiate immune responses<sup>2</sup>. In addition, it stores ~ 10% of the body's total red blood cells (RBC)<sup>3</sup>, which through contraction of smooth muscles<sup>4</sup>, can be expelled into the systemic circulation<sup>5–9</sup>, thereby enhancing oxygen carrying-capacity. There is great individual variation associated with splenic size, both in measures of specific diameters (maximal length, width and thickness)<sup>10</sup> and in resting volume, ranging from 58 to 375 mL in healthy individuals<sup>11–13</sup>.

Determination of normative splenic volumes is particularly important in medical work, to accurately diagnose, for example, splenomegaly<sup>13</sup> and various hematological diseases. In such clinical settings, helical computed tomography (CT) and magnetic resonance (MR) imaging are the gold standard methods<sup>14,15</sup>. However, these methods are also stationary and expensive. On the other hand, ultrasonography is a useful non-invasive, portable method, which is applicable in both laboratory settings and during fieldwork contexts. Ultrasonic imaging is particularly suited to experiments involving various stressors, such as maximal exercise<sup>3,16</sup>, voluntary apnea<sup>8,17</sup>, normobaric hypoxia<sup>18</sup> or hypobaric hypoxia<sup>19</sup>, which often includes a time-constrained protocol where windows for collection of splenic measurements are short. The length of the spleen display a strong correlation with splenic volumes from measures of cadavers<sup>20</sup> and with CT scan measurements<sup>14,21</sup>.

The first model to compute splenic volume measures based on ultrasonography was designed by Koga<sup>22,23</sup>, who formulated an equation based on measures of maximal splenic length and thickness yielding a measure of sectional area. More recent research, however, uses computation of maximal splenic length, width and thickness, to estimate the volume, whereby a prolate ellipsoid formula is considered the general convention<sup>11,13,14,24</sup>. Nonetheless, other formulas exist, one, in particular, was designed to account for individual differences in splenic diameters, which could plausibly result in varying splenic volumes. This particular formula has been broadly used in physiological laboratory research, involving apneic diving<sup>8,25,26</sup>, normobaric hypoxia exposure<sup>18,27–29</sup> and during maximal exercise<sup>16</sup>. Additionally, it has been used during field-based contexts involving hypobaric hypoxia and exercise<sup>17,19,30,31</sup>. However, the reliability of ultrasonography-derived splenic volumes between repeated measures using this particular formula is to date unknown.

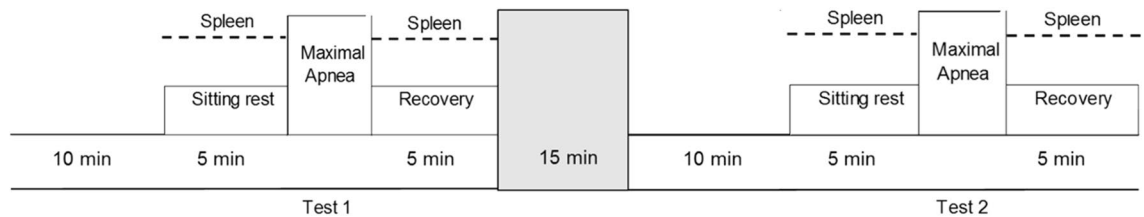
Determination of reliability of splenic volume assessed by ultrasound is important in physiological research, wherein the focus is towards assessment of the magnitude of a response induced by a specific stressor, e.g. hypoxia, apnea or maximal exercise. Therefore, the ability to replicate measurements consistently over time on repeated measurements, with the highest possible accuracy is important to be able to draw valid conclusions. To the best of our knowledge, research assessing the reliability of splenic volumes by ultrasonography is limited. Hosey et al.<sup>10</sup> reported moderate to good intra-rater reliability and good inter-rater reliability between repeated tests, however not on actual splenic volume measures or in response to apnea. Accordingly, by adopting a test–retest design, we firstly aimed to determine the absolute and relative reliability of splenic volume, assessed by ultrasonography at baseline and immediately after maximal voluntary apnea. Secondly, we aimed to determine the reliability of different calculation methods that are used to establish a baseline splenic volume, and thirdly, it was an aim to compare reliability measures of the Pilström formula with two other widely accepted formulas for splenic volume calculations.

## Methods

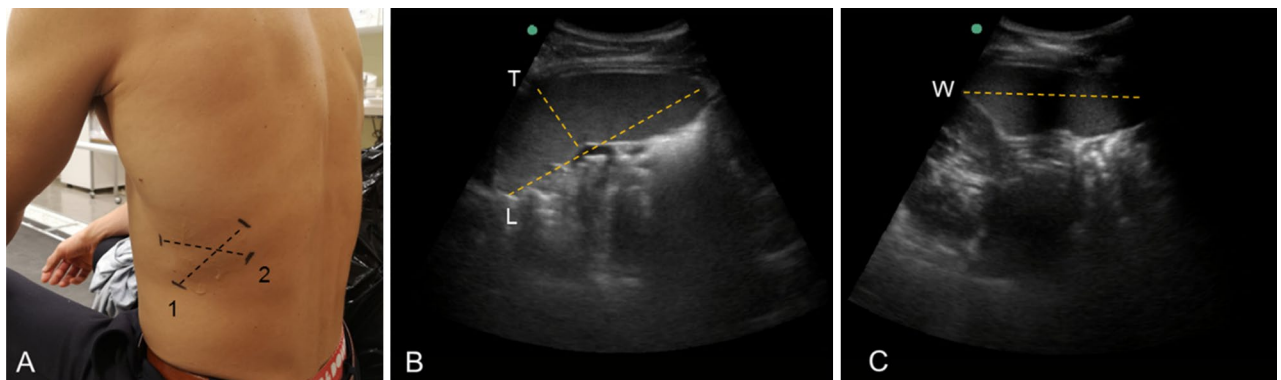
**Participants recruitment.** 12 healthy participants, inexperienced in apneic diving, volunteered for the investigation (six females and six males; age:  $30 \pm 6$  yrs; height:  $170 \pm 6$  cm; body mass:  $70 \pm 11$  kg; body mass index:  $24.4 \pm 3$  kg/m<sup>2</sup>). The participants were recruited via convenience sample from the department of health science at Mid Sweden University, either by answering official adverts at the department or by word a mouth. Sample size considerations were based on a required power of 0.8 and an alpha level fixed at 0.05 yielding a required sample size of minimum 10 participants in order to detect an intraclass correlation coefficient (ICC) of 0.8 and higher, given two observations per participant<sup>32</sup>. These considerations were made on the grounds that we expected splenic volume to exhibit low variations between tests, which we have found in previous pilot attempts in our laboratory, as such with lower variation a smaller sample would typically be required to detect higher ICC. Only one rater collected the measurements.

**Study overview.** The experiment comprised a test–retest design, consisting of two apnea tests (test 1 and test 2, Fig. 1), which were separated by 15 min. On termination of test 1, all equipment was detached and the ultrasound gel cleaned off, after which the participants were asked to stand and leave the laboratory while walking around. On return to the laboratory, test 2 commenced in the same manner as test 1. This design was chosen, aiming to exclude any factors that may result in temporary splenic changes (i.e. food intake, exercise, fasting, low O<sub>2</sub> saturation or infections), and possibly reduce unwanted variation. Participants were instructed to abstain from alcohol for at least 24 h before testing and from caffeine or tobacco on the day of the trial before testing. After height (cm) and body mass (kg) were measured (Seca 764, Hamburg, Germany), participants rested in a supine position for 10 min.

**Apnea tests.** The apnea test consisted of one static apnea of maximal voluntary duration in air, which was performed in a sitting position and preceded by spontaneous breathing that started after a coached inspiration on the request to take a deep, but not maximal breath. This technique generally results in inhalation of approximately 85% vital capacity<sup>33</sup>. Participants were notified when two and one minutes remained to initiate the breath-hold, when 30 s remained they were handed a nose clip and when 10 s remained a second-by-second countdown started. Participants were asked to refrain from hyperventilation before the apnea. If SpO<sub>2</sub> (peripheral oxygen saturation) fell below 65%, the participants were told to terminate the apnea and resume breathing,



**Figure 1.** Experimental protocol for test 1 and test 2. Both tests were identical and started with 10 min of lying (supine) rest, after which participants were moved to a sitting 5-min rest, at which splenic measurements were collected each minute for baseline (resting) splenic volume assessment. Thereafter, the maximal voluntary apnea was performed and immediately on termination, a 5-min resting recovery period followed, wherein splenic measurements were collected for apnea-induced volume assessments. Test 1 and test 2 were separated by 15 min, during which the participants were cleared of all testing equipment and asked to leave the laboratory, to return when test 2 should start.



**Figure 2.** Determination of splenic size in resting state with ultrasonography, which was collected on the dorsal side of the body (A). Representative ultrasound images are shown to illustrate the measurement of maximal splenic length (L) and thickness (T) in the longitudinal plane (B) and line 1 in (A) and width (W) in the transverse plane (C) and line 2 in (A).

to avoid the risk of hypoxic syncope. Upon termination of apnea, the nose clip was removed, and the participants were instructed to resume normal breathing.

**Splenic ultrasound measurements.** Participants were seated on a chair while the spleen site was identified on the dorsal side of the body (Fig. 2). The splenic size was measured via ultrasonic imaging (M-Turbo Ultrasound system, FUJIFILM Sonosite Inc, Bothell, WA USA) with the probe: C60x/5–2 MHz (SonoSite Inc, Bothell, WA, USA), by an experienced sonographer (PH) and measurements were collected each minute for five minutes during the rest period before the apnea, and immediately after termination of the apnea, during which measurements were collected each minute until five minutes after the apnea (Fig. 1). Two pictures were recorded each minute for determination of the three-axial maximal diameters of the spleen: length, thickness and width, and analyzed in real-time from still images (Fig. 2).

**Analysis.** The three-axial maximal diameters of splenic length (L), thickness (T) and width (W) were then used to calculate splenic volume according to the following equations:

$$\text{Pilström formula (1)} : V_{\text{spleen}} = (L\pi(WT - T^2))/3.$$

The equation is based on the observed average shape of the spleen detected on the ultrasonic images<sup>8</sup>, and have been used elsewhere in similar experiments involving splenic volume assessments<sup>17–19,25,30</sup>.

$$\text{Prolate ellipsoid formula(2)} : V_{\text{spleen}} = (L \times T \times W \times 0.524)$$

The equation describes rotation about the ellipses major axis, resulting in a prolate, which has been broadly used in physiological and clinical research<sup>11,13,14,24</sup>.

$$\text{Koga formula (3)} : V_{\text{spleen}} = (7.53S - 77.56)$$

This equation uses the sectional area (S), to estimate a representative volume measure<sup>23</sup>, which includes measurements of maximal splenic length and thickness<sup>22</sup>.

The standard approach for establishing a baseline splenic volume in this study uses the individual measurements obtained from the five minutes before the apnea, by averaging the ‘two consecutive maximal measurement’

	Test 1	Test 2	<i>p</i> value	CV (%)	TE (mL)	ICC	95% CI
<b>Pilström formula (1)</b>							
Two consecutive maximal measures (mL)	157 ± 39	156 ± 34	.651	2.98 ± 0.1	6.4	.970***	.900 to .991
Mean of all measures (mL)	157 ± 35	153 ± 34	.068	2.37 ± 0.1	4.5	.979***	.915 to .994
One single measure (mL)	161 ± 38	155 ± 37	.308	6.82 ± 0.6	14.3	.850***	.577 to .954
<b>Prolate ellipsoid formula (2)</b>							
Two consecutive maximal measures (mL)	154 ± 37	144 ± 43	.122	5.47 ± 0.3	8.6	.942***	.808 to .983
Mean of all measures (mL)	149 ± 36	148 ± 43	.722	4.27 ± 0.2	8.4	.954***	.850 to .987
One single measure (mL)	158 ± 43	149 ± 42	.133	7.52 ± 0.5	13.8	.879***	.619 to .962
<b>Koga formula (3)</b>							
Two consecutive maximal measures (mL)	142 ± 38	140 ± 59	.845	9.72 ± 1.4	17.6	.859***	.580 to .958
Mean of all measures (mL)	142 ± 39	142 ± 49	.997	7.07 ± 0.5	13.8	.911***	.718 to .973
One single measure (mL)	152 ± 49	143 ± 48	.241	9.25 ± 1.0	15.6	.892***	.681 to .967

**Table 1.** Comparisons of three different methods of establishing baseline splenic volume (Mean ± SD) between repeated tests using different formulas with corresponding reliability measures. \*\*\* $p < .001$ .

values. This particular method was implemented to limit any influence of variability that may occur due to pulsatile changes in splenic volume and minor changes in probe placement<sup>17,25</sup>. In addition, to compare reliability measures of different approaches of establishing baseline volume, a ‘mean of all measures’ were also computed, simply as an average of all splenic measures obtained from the five minutes before the apnea and ‘one single measure’. The single measure was taken as the value one minute before the apnea, to account for possible anticipatory-induced splenic contractions.

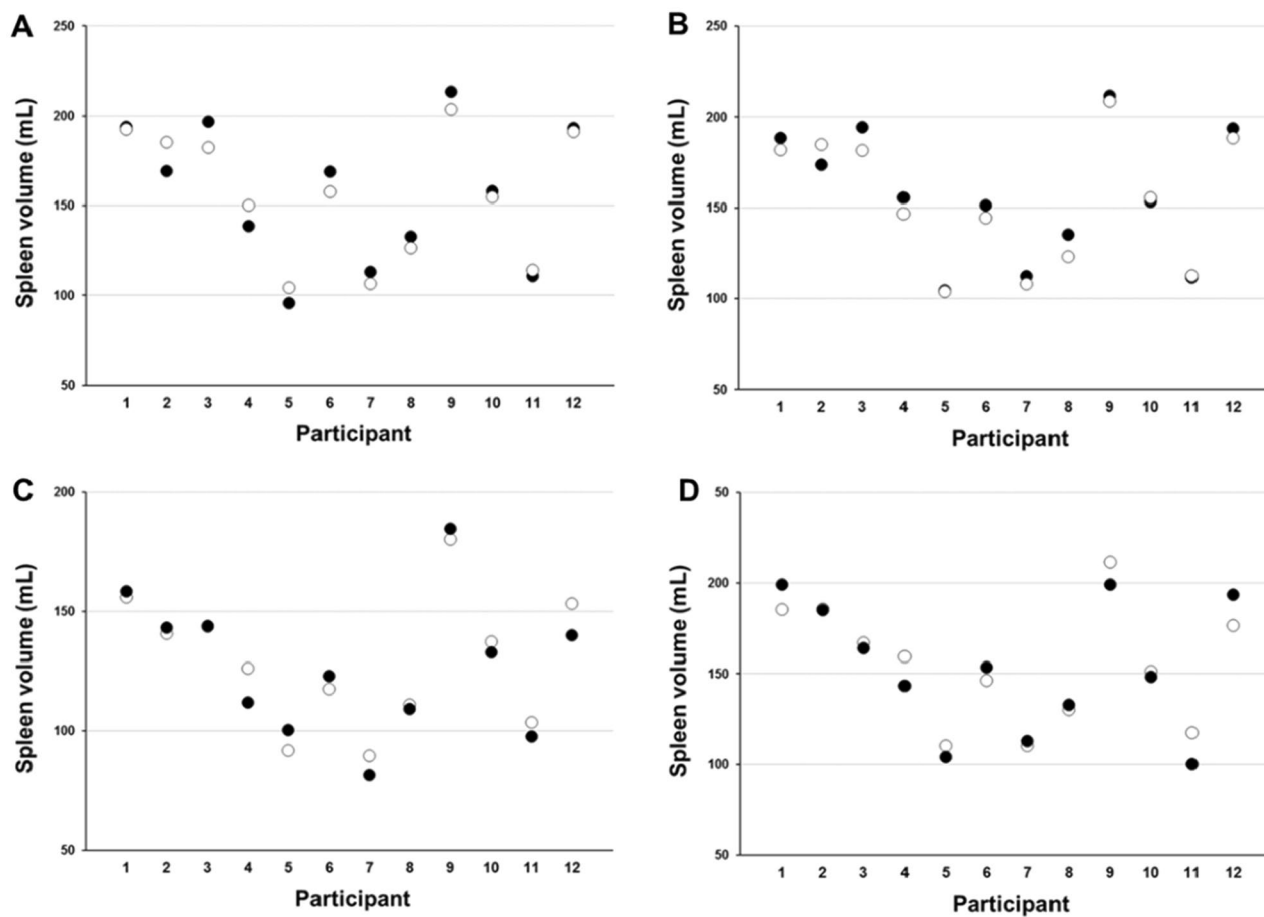
**Statistical analysis.** Data are presented as mean ± standard deviation (SD) and were analyzed using IBM SPSS 24.0 for Windows (SPSS Inc., Chicago, IL). Shapiro–Wilks and Kolmogorov–Smirnov tests ( $p > .05$ ) were used to assess if study variables were normally distributed. Paired samples *t*-tests were used to assess the difference in splenic volume between test 1 and test 2, as well as between formulas. To assess the relative reliability of splenic volumes between test 1 and test 2, and between formulas, ICCs were used, designed as a two-way mixed effect model with absolute agreement, which was considered appropriate to assess reliability with repeated measures including one rater<sup>34</sup>. The level of reliability was determined based on ICC: ICC less than 0.5 was considered poor reliability, ICC between 0.5 and 0.75 indicated moderate reliability, ICC values between 0.75 and 0.9 indicated good reliability and ICCs greater than 0.9 indicated excellent reliability<sup>34</sup>. ICCs were presented with 95% confidence intervals (95% CI). The coefficient of variation (CV) was calculated as the SD of the individual mean value of test 1 and test 2, after which the square root of the mean CV squared was calculated. Typical error (TE) was calculated as the SD of the difference between test 1 and test 2 divided by the square root. To further assess the agreement of splenic volumes, calculated by three different formula, Bland and Altman plots with associated limits of agreement (95% CI from  $-1.96$  SD to  $+1.96$  SD) was used as described by Giavarina<sup>35</sup>.

**Ethics approval.** The study abided by the regional ethical review board in Umeå, Sweden (reference: 2018-46-31M) and the Declaration of Helsinki. In addition, all measurements were non-invasive, thereby the experiment design did not expose the participants to any increased risks.

## Results

Baseline splenic volume displayed high reliability between tests for ‘two consecutive maximal measures’ (ICC = .970,  $p < .001$ , Table 1 and Fig. 3A) and ‘mean of all measures’ (ICC = .979,  $p < .001$ , Table 1 and Fig. 3B), and reliability was reduced for ‘one single measure’ (ICC = .850,  $p < .001$ , Table 1). Apnea induced-splenic volume also displayed high reliability for ‘two consecutive maximal measures’ (ICC = .967, 95% CI [.894 to .990],  $p < .001$ , Fig. 3C), with corresponding CV of  $3.14 \pm 3.1\%$  and TE of 4.9 mL. Splenic volume was significantly reduced immediately after apnea in both tests ( $p < .001$ ), to  $127 \pm 29$  mL ( $-18.0\%$ ) in test 1 and  $129 \pm 28$  mL ( $-16.5\%$ ) in test 2 (mean difference of 2.0 mL, 95% CI [ $-6.6$  to  $2.6$ ],  $p = .359$ ). After five minutes of recovery, splenic volume had reached baseline values in both tests ( $p > .05$ ), to  $153 \pm 36$  mL in test 1 and  $154 \pm 33$  mL in test 2 (mean difference of 1.2 mL, 95% CI [ $-8.1$  to  $5.6$ ],  $p = .706$ ; ICC = .954, 95% CI [.849 to .986],  $p < .001$ , Fig. 3D) and a corresponding CV of  $3.57 \pm 0.2\%$  and a TE of 7.3 mL.

Formula (1) displayed similar levels of reliability compared to formula (2) for all volume calculation methods, whereas formula (3) displayed lower reliability for each calculation method (Table 1). The reliability between each formula is reported in Tables 2 and 3. The agreement between formula (1) and formula (3) as assessed by Bland and Altman plots, displayed a bias (systemic measurement error) of  $-14.8$  mL (95% CI [ $5.4$  to  $-34.9$ ]), with the upper 95% limits of agreement of 47.3 mL and the lower 95% limits of agreements of  $-76.9$  mL (Fig. 4A) for ‘two consecutive maximal measures’ and a bias of  $-15.2$  mL (95% CI [ $2.1$  to  $-32.4$ ]), with the upper 95% limits of agreement of 38.1 mL and the lower 95% limits of agreements of  $-68.4$  mL (Fig. 4B) for ‘mean of all measures’. The agreement between formula (1) and formula (2) displayed a bias of  $-3.1$  mL (95% CI [ $11.1$  to  $-17.3$ ]), with the upper 95% limits of agreement of 40.7 mL and the lower 95% limits of agreements of  $-46.9$  mL (Fig. 4C) for



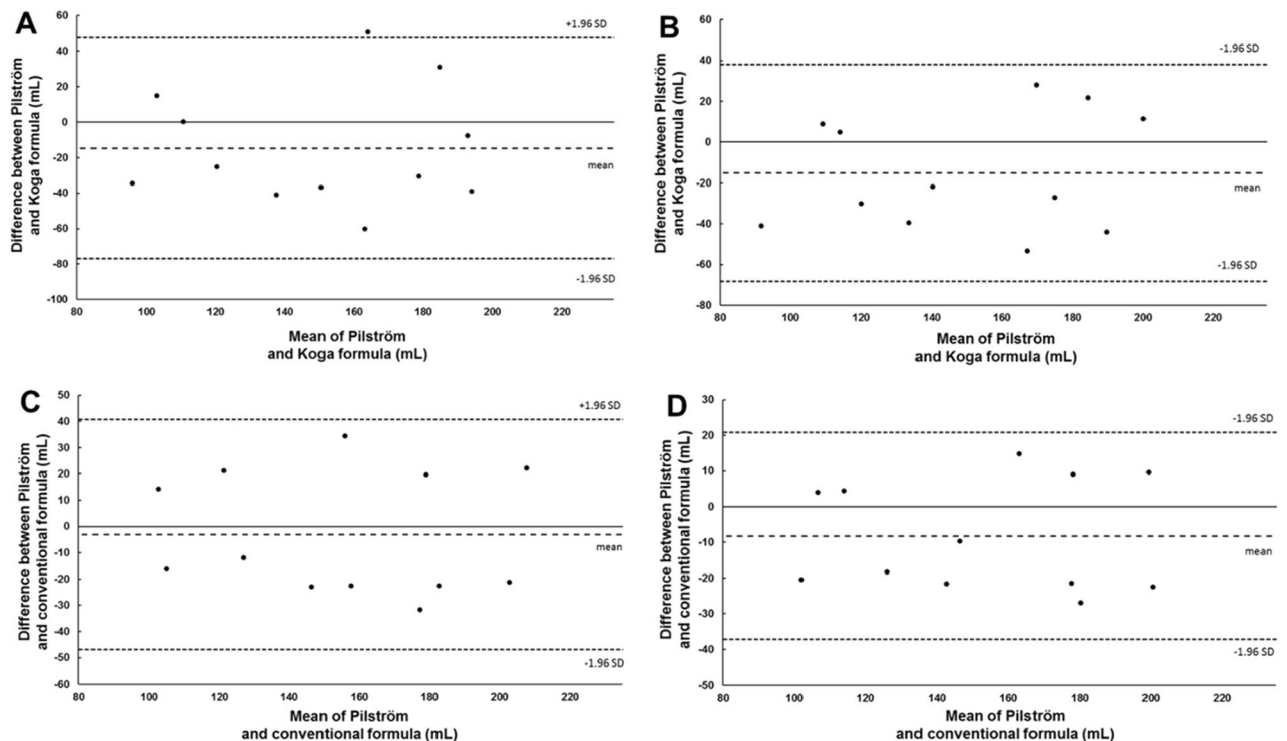
**Figure 3.** Individual splenic volumes for test 1 (black dots) and test 2 (white dots) for baseline (resting) splenic volume as calculated by ‘two consecutive maximal measures’ (mL; **A**), baseline splenic volume as calculated by ‘mean of all measures’ (mL; **B**), apnea-induced splenic volume and (mL; **C**) after five minutes of recovery (mL; **D**) for each participant using formula (1) (n = 12).

	Pilström formula (1)	Koga formula (3)	p value	CV (%)	TE (mL)	ICC	95% CI
Two consecutive maximal measures (mL)	157 ± 39	142 ± 38	.150	11.83 ± 1.1	22.4	.618**	.138 to .870
Mean of all measures (mL)	157 ± 36	142 ± 39	.091	11.24 ± 1.4	19.9	.693**	.242 to .900
One single measure (mL)	161 ± 38	152 ± 49	.377	12.59 ± 1.3	15.2	.671**	.205 to .892

**Table 2.** Comparisons of different methods of establishing baseline splenic volume (Mean ± SD) between Pilström and Koga formulas with corresponding reliability measures. \*\*p < .01. Values refer to comparisons of test 1.

	Pilström formula (1)	Prolate ellipsoid formula (2)	p value	CV (%)	TE (mL)	ICC	95% CI
Two consecutive maximal measures (mL)	157 ± 39	154 ± 37	.654	7.33 ± 0.3	15.8	.818***	.485 to .944
Mean of all measures (mL)	157 ± 36	149 ± 36	.092	5.67 ± 0.3	10.4	.892***	.652 to .968
One single measure (mL)	161 ± 38	158 ± 43	.586	6.96 ± 0.3	15.2	.857***	.583 to .957

**Table 3.** Comparisons of different methods of establishing baseline splenic volume (Mean ± SD) between Pilström and Conventional formulas with corresponding reliability measures. \*\*\*p < .001. Values refer to comparisons of test 1.



**Figure 4.** Bland and Altman plot for baseline splenic volume measures. The upper panel displays splenic volume difference (mL; Y-axis) between, and mean splenic volume (mL; X-axis) of formula (1) and formula (3) using ‘two consecutive maximal measures’ (A) and ‘mean of all measures’ (B). The lower panel displays splenic volume difference (mL; Y-axis) between, and mean splenic volume (mL; X-axis) of formula (1) and the formula (2) using ‘two consecutive maximal measures’ (C) and ‘mean of all measures’ (D). The bias is represented (middle dotted line) along with zero difference (solid line) with the representation of the upper and lower limits of agreements (small dotted lines), from  $-1.96$  SD to  $+1.96$  SD ( $n = 12$ ).

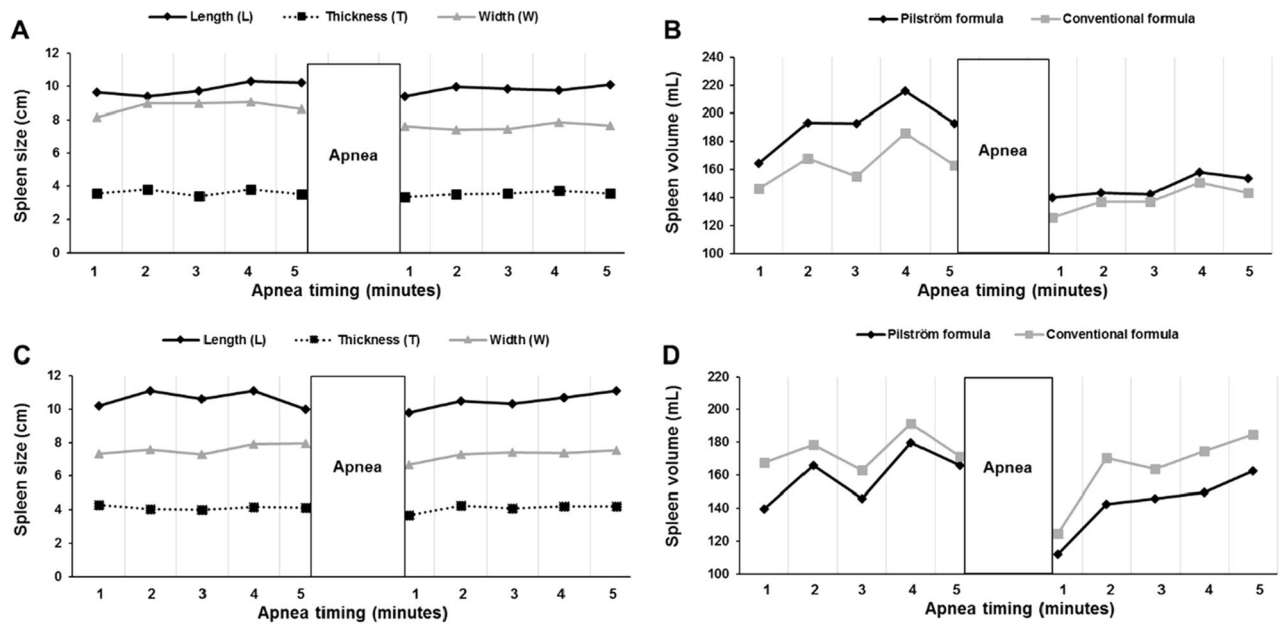
‘two consecutive maximal measures’ and a bias of  $-8.2$  mL (95% CI [1.1 to  $-17.6$ ]), with the upper 95% limits of agreement of 20.7 mL and the lower 95% limits of agreements of  $-37.2$  mL (Fig. 4D) for ‘mean of all measures’.

The two-volume measures which were used to determine baseline splenic volume through the ‘two consecutive maximal measures’ method displayed CV of  $2.23 \pm 0.04\%$  (TE = 4.8 mL) and  $1.14 \pm 0.01\%$  (TE = 2.0 mL) for test 1 and test 2 respectively. Figure 5 illustrates the individual variation in splenic volume values in response to apnea between formula (1) and formula (2), as a function of varying splenic diameters.

## Discussion

This study aimed to determine the test–retest reliability of splenic volume assessed by ultrasonography at baseline and in response to maximal voluntary apnea and to compare reliability measurements between different calculation methods and between different splenic volume formulas. The main findings of this report are as follows: (a) baseline splenic volume and apnea-induced volume indicated excellent test–retest reliability between repeated tests, (b) assessment of different baseline volume calculation methods showed that ‘two consecutive maximal measures’ and ‘mean of all measures’ displayed the highest reliability between repeated tests and (c) Pilsström formula (1) displayed slightly higher reliability between repeated tests compared to the Koga formula (3), but similar reliability compared to the prolate ellipsoide formula (2) and high reliability and agreement between the two formulas.

Our data indicate high accuracy and consistency in splenic measures during an apnea test of maximal voluntary effort, assessed by ultrasonography between two identical test situations. Specifically, ICCs of resting splenic volume displayed excellent intra-rater reliability, showing that ultrasonography-derived splenic volume at rest is consistent on repeated measurements, highlighted by strong agreement and correlation of measurements between tests, as calculated by ‘two consecutive maximal measurements’ and ‘mean of all measures’. This is in agreement with previous reports by Hosey et al.<sup>10</sup>, who also found high intra-rater reliability. However, their assessments are only based on the length and width of the spleen, as opposed to a calculated splenic volume. Further, the CV% of the baseline volume was also notably low, indicating low within-subject variability between test occasions. Expressed in practical terms, the splenic volume will vary with 4.5 to 6.4 mL on repeated measurements, using ‘two consecutive maximal measures’ and ‘mean of all measures’ using formula (1), which is notably low considering normative splenic volumes<sup>11,13</sup>. On the other hand, ‘one single measure’ displayed lower consistency between test occasions, as shown by moderate to good intra-rater reliability and higher CV%, with TE of  $\sim 15$  mL, for all formulas. This increased variability of the ‘one single measure’ method, that only takes one value from the five minutes before the apnea as a representative baseline splenic volume, seems logical, as the spleen is highly



**Figure 5.** Individual variation in splenic volume as a function of variations in splenic diameters. Panel (A) and (C) display the size of each splenic diameter (length, thickness and width; cm) and panels (B) and (D) display splenic volume (mL), calculated by formula (1) (black line) and formula (2) (grey line) from the measures in panel (A) and (C), for two participants [nr 1 (A, B) and nr 2 (C, D)]. The apnea test started with five minutes of sitting rest, during which splenic measurements (A, C) were collected each minute for determination of baseline volume (B, D), following one maximal voluntary apnea. Thereafter, an additional five minutes of rest followed, during which splenic measures (A, C) were collected each minute, for apnea-induced contraction and recovery volume (B, D).

vascularized with a continuous pulsating blood flow<sup>1</sup> resulting in high-frequency changes in volume throughout the rest period. Consequently, our data indicate that ‘two consecutive maximal measures’ and ‘mean of all measures’ are more reliable and consistent compared to ‘one single measure’ and therefore represent a more appropriate method to establish a resting baseline splenic volume. Which of the two methods to employ, should be based upon individual volume variations during the baseline splenic measurement period. We found a CV of between 1.1 to 2.2% between the two-volume measures used as the ‘two consecutive maximal measures’ method, therefore we suggest that a CV of approximately 3% or less, should be sought after if this particular method should be implemented, as a means to increase reliability. The rationale for the ‘two consecutive maximal measures’ method is that it may limit variation brought about by either the pulsatile changes in volume due to continuous blood flow (biological variation) and/or minor differences in probe placements (measurement error) between collections. Consequently, two similar volume measurements that come in succession with a one-minute interval would likely reflect a more valid baseline volume, effectively excluding values due to pulsatile blood flow or varying probe placements. However, if the volume variation is too great and the two-volume measurements no longer are similar to one another, i.e. with increased variation, this assumption is violated and the ‘mean of all measures’ method should preferably be employed.

In addition, the apnea-induced splenic volume exhibited similar, excellent intra-rater reliability and low within-subject variability, indicating that this consistency of splenic volumes assessed by ultrasound, is similar also in response to a specific stressor, suggesting that the magnitude of splenic contraction induced by hypoxia is consistent on repeated measurements. Therefore, splenic volumes obtained by ultrasonic imaging have practical applications during apnea, as a method detecting splenic volume change consistently between repeated measurements. In contrast to clinical research, which is more inclined towards the determination of normative splenic volumes to accurately diagnose various symptoms and disorders. Physiological research is predominantly interested in the assessment of magnitudes of responses induced by a stressor or by measurements between groups. Consequently, it is vital to, with high accuracy, replicate measurements consistently between repeated measurements and between groups. Our data indicate that splenic volumes recorded by ultrasonography are highly reliable with low variation, which has practical applications during physiological laboratory and field-based investigations involving various stressors, including hypoxia.

The splenic length measured by ultrasound is generally strongly correlated with the actual splenic volume of cadavers<sup>20</sup> and with measurements derived by CT<sup>14,21</sup>, indicating that splenic volumes recorded by ultrasonography are valid representations of the actual splenic size. Further, splenic volume change in response to apnea has been assessed by MR imaging, which, in line with splenic volume change assessed by ultrasound, also displays a significant apnea-induced splenic contraction<sup>15</sup>. The conventional method to calculate splenic volume from the three diameters is typically formula (2), which applies a function of 0.524<sup>11,13</sup> and have also, indirectly, been validated<sup>14</sup>. Koga<sup>23</sup> was the first study to calculate splenic volume using ultrasonography, however, this formula is limited to only include two splenic dimensions and estimating splenic volume based on the sectional area,

without any information regarding splenic width, typically measured in the transverse plane. Therefore, it is highly plausible that formula (3) over or underestimates splenic volumes of individuals with particularly atypical splenic shapes. Our data indicate that formula (3) has reduced consistency between repeated tests, pertaining to both CV% and ICC, for each baseline calculation method compared to both formula (1) and formula (2). One likely reason for this slightly lower reliability of formula (3) could likely be ascribed to the fact that it does not include triaxial measurements and thereby are less representative of individual variations in splenic shapes.

Formula (1)<sup>8,17,30</sup>, was designed in response to individual variation in splenic diameters<sup>10,11</sup>, which potentially could result in inconsistent splenic volumes. Our data highlights minor differences in splenic volume between formulas. By viewing Fig. 4C,D it is quite clear that the differences between formula (1) and formula (2) are small within each participant for both ‘two consecutive maximal measures’ and ‘mean of all measures’, denoted by a small systematic measurement error (bias), wherein none of the participants in the sample falls outside the limits of agreements, displaying high agreement. Second, it is also clear, that the calculated splenic volumes for both formulas are not consistently higher or lower than the other. Thus, one of the formulas does not overestimate or underestimate volumes derived by the other formula. Instead, different splenic volumes likely depend on specific measures of individual splenic diameters. This may be further substantiated by Fig. 5 that shows variations in splenic volumes between formulas as a function of varying diameters, wherein splenic width likely leads to a greater degree of volume variations. This may suggest that formula (1) may account for individual differences in splenic dimensions in a dynamic way and therefore reflect varying splenic shapes. Although the agreement between formula (3) and formula (1) was good at best, with a systematic measurement error within 15 mL, it was lower compared to the agreement between the formulas including triaxial measurements.

Nevertheless, even though the differences in splenic volumes between formulas are relatively small, displaying good to excellent ICCs and minor CV%, which indicates high reliability, the accuracy of the measurement error, as assessed by 95% limits of agreements, was relatively low, particularly between formula (1) and formula (3). This low accuracy of the measurement error can potentially result in substantial differences in splenic volumes between formulas and should be taken into account when comparing splenic volumes across studies using different formulas. However, as all formulas display high ICCs and low within-subject variation in baseline splenic volume across repeated measurements, comparisons between splenic response magnitudes are highly relevant between formulas, but comparisons of absolute splenic volume values should preferably be avoided.

## Conclusion

We conclude that both baseline splenic volume and the apnea-induced volume assessed by ultrasonography, has excellent reliability between repeated tests for both the Pilström formula and the prolate ellipsoid formula, suggesting that ultrasonographic measurements of the spleen have practical application during laboratory and field-based research involving apnea and hypoxia, as a reliable method detecting splenic volume change, consistently between repeated measurements. Additionally, to determine an accurate baseline splenic volume, our data highlights two methods: (1) averaging the two consecutive maximal volume measurements during a five-minute resting period and (2) obtaining an average of all volume measurements during a five-minute resting period. Determination of which of the two methods to implement should preferably be assessed based on individual volume variation between the selected two maximal values. The Pilström formula exhibited slightly higher reliability compared to the Koga formula, but similar reliability and variability compared to the prolate ellipsoid formula, with a small systematic measurement error showing high agreement, suggesting that the Pilström formula may be particularly useful to account for individual differences in splenic dimensions. However, the accuracy of the measurement error was low, consequently, comparisons of absolute splenic volumes between studies using different formulas should be avoided.

## Data availability

All data generated or analysed during this study are included in this published article.

Received: 12 January 2022; Accepted: 31 October 2022

Published online: 08 November 2022

## References

1. Cesta, M. F. Normal structure, function, and histology of the spleen. *Toxicol. Pathol.* **34**, 455–465 (2006).
2. Mebius, R. E. & Kraal, G. Structure and function of the spleen. *Nat. Rev. Immunol.* **5**, 606–616 (2005).
3. Stewart, I. B. & McKenzie, D. C. The human spleen during physiological stress. *Sports Med.* **32**, 361–369 (2002).
4. Pinkus, G., Warhol, M., O’connor, E., Etheridge, C. & Fujiwara, K. Immunohistochemical localization of smooth muscle myosin in human spleen, lymph node, and other lymphoid tissues. Unique staining patterns in splenic white pulp and sinuses, lymphoid follicles, and certain vasculature, with ultrastructural correlations. *Am. J. Pathol.* **123**, 440 (1986).
5. Bakovic, D. *et al.* Spleen volume and blood flow response to repeated breath-hold apneas. *J. Appl. Physiol.* **95**, 1460–1466 (2003).
6. Laub, M. *et al.* Spleen emptying and venous hematocrit in humans during exercise. *J. Appl. Physiol.* **74**, 1024–1026 (1993).
7. Schagatay, E., Andersson, J., Hallén, M. & Pålsson, B. Selected contribution: Role of spleen emptying in prolonging apneas in humans. *J. Appl. Physiol.* **90**, 1623–1629 (2001).
8. Schagatay, E., Haughey, H. & Reimers, J. Speed of spleen volume changes evoked by serial apneas. *Eur. J. Appl. Physiol.* **93**, 447–452 (2005).
9. Schagatay, E., Andersson, J. & Nielsen, B. Hematological response and diving response during apnea and apnea with face immersion. *Eur. J. Appl. Physiol.* **101**, 125–132 (2007).
10. Hosey, R. G. *et al.* Ultrasound assessment of spleen size in collegiate athletes. *Br. J. Sports Med.* **40**, 251–254 (2006).
11. De Odorico, I. *et al.* Normal splenic volumes estimated using three-dimensional ultrasonography. *J. Ultrasound Med.* **18**, 231–236 (1999).
12. Prassopoulos, P., Daskalogiannaki, M., Raissaki, M., Hatjidakis, A. & Gourtsoyiannis, N. Determination of normal splenic volume on computed tomography in relation to age, gender and body habitus. *Eur. Radiol.* **7**, 246–248 (1997).



13. Chow, K. U., Luxembourg, B., Seifried, E. & Bonig, H. Spleen size is significantly influenced by body height and sex: Establishment of normal values for spleen size at US with a cohort of 1200 healthy individuals. *Radiology* **279**, 306–313 (2016).
14. Yetter, E. M., Acosta, K. B., Olson, M. C. & Blundell, K. Estimating splenic volume: Sonographic measurements correlated with helical CT determination. *Am. J. Roentgenol.* **181**, 1615–1620 (2003).
15. Inoue, Y., Nakajima, A., Mizukami, S. & Hata, H. Effect of breath holding on spleen volume measured by magnetic resonance imaging. *PLoS ONE* **8**, e68670. <https://doi.org/10.1371/journal.pone.0068670> (2013).
16. Holmström, P., Karlsson, Ö., Lindblom, H., McGawley, K. & Schagatay, E. Enhanced splenic contraction in elite endurance athletes. *J. Appl. Physiol.* **131**, 474–486 (2021).
17. Holmström, P., Mulder, E., Lodin-Sundström, A., Limbu, P. & Schagatay, E. The magnitude of diving bradycardia during apnea at low-altitude reveals tolerance to high altitude hypoxia. *Front. Physiol.* <https://doi.org/10.3389/fphys.2019.01075> (2019).
18. Lodin-Sundström, A. & Schagatay, E. Spleen contraction during 20 min normobaric hypoxia and 2 min apnea in humans. *Aviat. Space Environ. Med.* **81**, 545–549 (2010).
19. Holmström, P. *et al.* The effects of incremental ascent to altitude on the diving response and splenic contraction during voluntary breath holding. *J. Exp. Physiol.* **106**, 160–174. <https://doi.org/10.1113/EP088571> (2020).
20. Loftus, W. K., Chow, L. T. & Metreweli, C. Sonographic measurement of splenic length: Correlation with measurement at autopsy. *J. Clin. Ultrasound* **27**, 71–74 (1999).
21. Lamb, P. *et al.* Spleen size: How well do linear ultrasound measurements correlate with three-dimensional CT volume assessments?. *Br. J. Radiol.* **75**, 573–577 (2002).
22. Koga, T. & Morikawa, Y. Ultrasonographic determination of the splenic size and its clinical usefulness in various liver diseases. *Radiology* **115**, 157–161 (1975).
23. Koga, T. Correlation between sectional area of the spleen by ultrasonic tomography and actual volume of the removed spleen. *J. Clin. Ultrasound* **7**, 119–120 (1979).
24. Purdy, G. M. *et al.* Spleen reactivity during incremental ascent to altitude. *J. Appl. Physiol.* **126**, 152–159 (2018).
25. Schagatay, E., Richardson, M. & Lodin-Sundström, A. Size matters: Spleen and lung volumes predict performance in human apneic divers. *Front. Physiol.* <https://doi.org/10.3389/fphys.2012.00173> (2012).
26. Elia, A., Barlow, M. J., Wilson, O. J. & O'Hara, J. P. Splenic responses to a series of repeated maximal static and dynamic apnoeas with whole-body immersion in water. *Exp. Physiol.* **106**, 338–349 (2021).
27. Richardson, M., Lodin, A., Reimers, J. & Schagatay, E. Short-term effects of normobaric hypoxia on the human spleen. *Eur. J. Appl. Physiol.* **104**, 395–399 (2008).
28. Pernett, F., Schagatay, F., Vildevi, C. & Schagatay, E. Spleen contraction during sudden eupneic hypoxia elevates hemoglobin concentration. *Front. Physiol.* **12**, 729123. <https://doi.org/10.3389/fphys.2021.729123> (2021).
29. Lodin-Sundström, A., Holmström, P., Ekstam, M., Söderberg, D. & Schagatay, E. Splenic contraction is enhanced by exercise at simulated high-altitude. *Eur J Appl Physiol* **121**, 1725–1732. <https://doi.org/10.1007/s00421-021-04637-0> (2021).
30. Egan, H., Lodin-Sundström, A., Schagatay, F. & Schagatay, E. The effect of climbing Mount Everest on spleen contraction and increase in hemoglobin concentration during breath holding and exercise. *High Alt. Med. Biol.* **15**, 52–57 (2014).
31. Schagatay, E., Lunde, A., Nilsson, S., Palm, O. & Lodin-Sundström, A. Spleen contraction elevates hemoglobin concentration at high altitude during rest and exercise. *Eur. J. Appl. Physiol.* **120**, 26–2704. <https://doi.org/10.1007/s00421-020-04471-w> (2020).
32. Bujang, M. A. & Baharum, N. A simplified guide to determination of sample size requirements for estimating the value of intraclass correlation coefficient: A review. *Arch. Orofac. Sci.* **12**, 1–11 (2017).
33. Schagatay, E. & Holm, B. Effects of water and ambient air temperatures on human diving bradycardia. *Eur. J. Appl. Physiol.* **73**, 1–6 (1996).
34. Koo, T. K. & Li, M. Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* **15**, 155–163 (2016).
35. Giavarina, D. Understanding bland altman analysis. *Biochemia Medica* **25**, 141–151 (2015).

## Acknowledgements

On behalf of all the authors, we give our utmost gratitude to all the volunteering participants, for their time and participation in this research investigation. We also thank Öyvind Karlsson, a Ph.D. student at Mid Sweden University for his establishment of the ethical application.

## Author contributions

P.H.: Research idea, general organization, data collection, data analysis and manuscript writing. F.P.: Laboratory organization, data collection, and critical review of the manuscript. E.S.: critical review and manuscript writing. All authors reviewed the manuscript. The authors affirm that human research participants provided informed consent to participate and for publication of the images in Fig. 2.

## Funding

Open access funding provided by Mid Sweden University.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-23384-6>.

**Correspondence** and requests for materials should be addressed to P.H.

**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) 2022