Real-time three-dimensional echocardiography for left atrial volume assessment in Thoroughbred racehorses: observer variability and comparison with two-dimensional echocardiography

Citation for published version:
Worsman, F, Miller, Z, Shaw, D, Blissitt, K & Keen, J 2020, 'Real-time three-dimensional echocardiography for left atrial volume assessment in Thoroughbred racehorses: observer variability and comparison with two-dimensional echocardiography', *Equine Veterinary Journal*. https://doi.org/10.1111/evj.13408

Digital Object Identifier (DOI):
10.1111/evj.13408

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Equine Veterinary Journal

DR. FRANCESCA WORSMAN (Orcid ID : 0000-0002-7273-9846)

DR. JOHN A KEEN (Orcid ID : 0000-0002-7862-5321)

Article type : General Article

Real-time three-dimensional echocardiography for left atrial volume assessment in Thoroughbred racehorses: observer variability and comparison with two-dimensional echocardiography

F. C. F. Worsman*, Z. Miller, D. J. Shaw, K. J. Blissitt and J. A. Keen

The Royal (Dick) School of Veterinary Studies, University of Edinburgh, Roslin, Midlothian EH25 9RG.

*Corresponding author email: fworsman@exseed.ed.ac.uk

Keywords: horse, equine left atrial volume, 3DE, intra-class correlation coefficient

Summary

Background: Left atrial size predicts cardiac morbidity and mortality in humans and dogs. Real-time three-dimensional echocardiography (3DE) may be reliable for assessing left atrial volume (LAV) in horses.

Objectives: To determine intra and inter-observer variability estimates of 3DE-LAV and compare it to that of 2DE-LAV estimates.

Study design: Method comparison.

Methods: 3DE datasets were obtained from 40 horses, then graded for quality, creating a final study population of 22 horses. The 3DE and 2DE maximum LAV (LAV_{max}) and minimum LAV (LAV_{min}) were measured, and left atrial emptying volume (LA EV) and left atrial ejection fraction (LA EF) calculated, from the same 3D dataset on four occasions using a) a semi-automatic surface recognition algorithm and b) a modified Simpson's method of discs. 3DE LAV measurements were repeated by a second observer.

Results: For 3DE, median LAV_{max} was 596cm^3 for observer one, and 852cm^3 for observer two, LAV_{min} was 373cm^3 for observer one and 533cm^3 for observer two. Low intra-observer measurement variation was observed for LAV_{max} and LAV_{min}, with horse-level intra-class correlation coefficients (ICC_{horse}) for both observers between 76-85% (horse added as random effect). The inter-observer ICC was 58% for LAV_{max} and 50% for LAV_{min} on averaged measurements (with observer added as random effect), indicating consistent differences between observers. While intra-observer variation was similar for 2DE LAV_{max} measurements, it was greater for LAV_{min} (ICC_{horse} = 67%). The inter-method ICC for 3DE vs 2DE was low at 14% for LAV_{max} and ~0% for LAV_{min}, indicating less-consistent differences with method.

Main limitations: Small study population, low observer number, use of different imaging modalities (fundamental frequency and octave harmonics).

Conclusions: 3DE assessment of LAV was reliable, suggesting suitability for longitudinal evaluation of clinical cases. Clinicians should be aware of differences in LAV measurements between observers. More defined measurement guidelines may improve repeatability.

Introduction

Accurate assessment of left atrial size is important and has been shown to predict morbidity and mortality in human and small animal veterinary cardiology [1, 2, 3, 4, 5]. In horses, accurate and repeatable measurement of the left atrium is of particular interest for quantifying volume overload and atrial systolic function in mitral valve regurgitation [6] and atrial fibrillation [7]. Left atrial size is routinely measured as part of the equine echocardiographic examination [6, 8], but measurements have been typically limited to linear dimensions from 2DE datasets, such as the left atrial diameter in a left or right parasternal long-axis view [9, 10] or the right parasternal short axis view. Linear dimensions can be misleading when assessing left atrial size, as the atrium has a complex three-dimensional topography [11] and, when enlarged, linear measurements may not reflect all dimensions of expansion [6]. More recently, 2DE area measurements have been described and used to calculate functional changes in left atrial size [12]. These images may also be used to estimate volume [12], but such techniques also rely on geometric assumptions. In addition, linear and area measurements by 2DE may be inaccurate owing to foreshortening of the image, which can be difficult to determine. Volume measurement using real-time three-dimensional echocardiography (3DE) makes no such geometric assumptions while mitigating against foreshortening and thus may provide a more accurate and repeatable technique for assessing left atrial volume [13,14,15,16] providing the whole atrium is included in the 3DE dataset. The aim of this study was to assess the repeatability and reproducibility of 3DE to assess left atrial volume estimates in horses, using a software algorithm designed for volume measurement of the left ventricle, and to compare this to estimated volumes calculated from 2DE measurements.

Materials and methods

3DE datasets were obtained from 40 healthy, athletic Thoroughbred horses each examined once. All horses were examined by cardiac auscultation and horses with murmurs higher than grade 2/6 were excluded from the study, as were horses with a heart rate greater than 48bpm. Images were obtained with an ultrasound scanner (Vivid E9[1]) using an active matrix 4D volume phased array transducer (4V-D[1]) at a frequency of 1.7/3.3 MHz with (*harmonic imaging*) and 2.5MHz without (*fundamental frequency)* octave harmonics. A single operator (obs_2) acquired all the images prospectively for the purpose of this study. A single-lead electrocardiogram (ECG) was recorded simultaneously. All images of the left atrium, single cycle multi-beat compilation of 4 cardiac cycles, were obtained from a modified standard right parasternal long-axis view, focusing on the left atrium. Significant time was spent acquiring good quality images ensuring the entire left atrium was incorporated with a low number of stitch artefacts whilst the horse had a low resting heart rate.

Observers analysed the studies offline, retrospectively with analysis software (EchoPac PC version 202[2]). The images were graded to select only good quality left atrial datasets for analysis (Supplementary Item 1). Images were included if the volumes per second was greater than 15 and all of the atrium was included within the frames. Consequently, by using the grading system, for each horse datasets with the best quality image and highest frame rate were chosen. The same multi-beat compiled cardiac cycle was used for both 2DE and 3DE measurements. Random-generated order, blinded measurements were obtained on 4 occasions by two observers, one internal medicine resident with further cardiology training (obs_1) and one

European boarded internal medicine Diplomate (*obs_2*) for 3DE; and a single observer (*obs_1*) for 2DE images. Maximum and minimum LAV were measured.

Maximum 3DE left atrial volume (3DE LAV_{max}) was determined to be at the last frame before the mitral valve opened and minimum 3DE left atrial volume (3DE LAV_{min}) was determined to be at the last frame before the mitral valve closed [17]. The 3DE LAV was measured using a semi-automated algorithm software (4D Auto LVQ in Manual Function; EchoPAC *v. 202*[2]), designed for assessing left ventricular volume. This software was used, because at the time of this study, software designed for 3DE atrial volume measurement was not available. The manual option of 4D Auto LVQ was chosen, because a pilot study (unpublished data) revealed that, the fully-automated measurement package (4D Auto LVQ in Auto Function; EchoPAC *v. 202*[2]) was unreliable at detecting the equine atrial wall and required significant endocardial marker adjustment.

The three left atrial slices (60° to the long axis), derived from the 3DE dataset, were re-aligned in the software to orientate the dorsal atrium uppermost, consistent with the apical orientation used to measure ventricular volume in human medicine. The long axis four chamber view was aligned first to allow the corresponding intersection line of all planes to be placed in the middle of the atrial cavity, crossing the theoretical apex (dorsal atrium) and the centre of the mitral valve opening in each view [18]. Aligning one plane automatically changed the other planes. The alignment of the other views was manually refined. These initial steps were completed for maximum atrial volume and did not need to be repeated to acquire minimum atrial volume which was taken to be at the end of ventricular diastole and is automatically detected by the software, the point at which it expects there to be maximum ventricular volume.

For maximum left atrial volume, the *Move ED*[2] button was therefore manually adjusted to the correct frame before the mitral valve opened (Supplementary Item 2). For minimum atrial volume the *Move ES*[2] button was similarly adjusted to the correct frame before the mitral valve closed. The subsequent steps were performed to acquire maximum left atrial volume and repeated to acquire minimum left atrial volume. At each time point, two electronic markers were placed on the endocardium at the hinge points either side of the mitral valve annulus and one was placed at the theoretical apex (dorsal aspect of the atrium) on all three 60° long axis slices. Positioning of the markers was simultaneously shown on the other long axis slices as well as one short axis slice for further guidance. The software then automatically generated a volume by endocardial border recognition in six atrial views. The LAV was then minimally adjusted manually to ensure accurate tracking and the final LAV was generated. The pulmonary veins and left atrial appendage were not included. The software then generated a dynamic surface rendered left atrial cast and the resulting volume-time plot (Figure 1) was examined to confirm that it demonstrated normal atrial filling and bi-phasic atrial emptying pattern.

For 2DE analysis, LAV was measured by a modified Simpson's method of discs (2DE Volume Biplane function EchoPAC v 202[2]) on the same single cycle-multi beat image. The atrium was measured in a single plane (corresponding to a right parasternal long axis 4 chamber image) for maximum volume by first positioning two markers on the endocardium either side of the mitral valve annulus then positioning a third marker bisecting the dorsal wall. The endocardial border was then traced by the operator using the software and, following manually marking the length of the chamber, a volume compromising multiple stacked discs was automatically generated. Manual adjustments after automated tracing were not performed in this technique. The method was repeated for measuring the minimum volume. The pulmonary veins and left atrial appendage were not included.

For both observers for 3DE and for obs_1 for 2DE, left atrial emptying volumes (LA EV) and left atrial ejection fractions (LA EF) were also calculated. The LA EV was calculated by subtracting the minimum LAV (LAV_{min}) from the maximum LAV (LAV_{max}), the LA EF was calculated by dividing the LA EV by the LAV_{max}.

Data analysis

All statistical analyses were carried out in R (v 4.0.0, © 2020 The R Foundation for Statistical Computing)[3]. All analyses were carried out on combined and separated fundamental frequency and harmonic imaging data. Repeatability was defined as the ability of the same single observer to obtain a same/similar result on a repeated measurement performed on the same sample. Reproducibility was defined as the ability of different observers to come up with a same/similar measurement. Intra-observer variability in both 3DE (in both observers, obs_1 and obs_2) and 2DE (obs_1) measurements for LAV_{max} and LAV_{min}, LA EV and LA EF were assessed via random effect linear models. Which horse the four repeated measures had come from for each observer/method, was entered as the random effect, and the associated horse-level intra-class correlation coefficients (ICCs) calculated (using the *lmer* and *sjPlot* packages[3]) (Figure 2). A high horse-associated ICC (ICC_{horse}) was taken to indicate low intra-observer variability i.e. the majority of the variability was attributable to the horse. For horse-associated ICCs intra-observer repeatability and reproducibility were determined to be excellent if >90%, good if >75%, moderate if between 74% and 50% and poor if <50% [19]. Also, as well as determining the standard deviation of the four repeated measures per horse, a repeatability coefficient, the absolute value below which the difference between the measurements will lie with 95% probability was determined. A Repeatability Coefficient (*RC*) was calculated as √2 x 1.96 x within-subject standard deviation [20].

Inter-observer variability in 3DE LAV_{max} and 3DE LAV_{min} were also assessed using ICC ($ICC_{inter\text{-}observer}$). The four repeated-measures per observer/method were averaged and random-effect linear models on mean 3DE LAV_{max} and 3DE LAV_{min}, 3DE LA EV and 3DE LA EF data, with observer added as the random effect were carried out (Figure 2). A high $ICC_{inter\text{-}observer}$ was taken to indicate high variability between observers i.e. the minority of the variability was attributable to the horse. Inter-method variability (3DE vs. 2DE) in obs_1 were also assessed via ICC ($ICC_{inter\text{-}method}$), with method added as the random effect and a high $ICC_{inter\text{-}method}$ was taken to indicate high variability in LAV between 3DE/2DE methods. Therefore, the $ICC_{inter\text{-}method}$ compares the correlation among 4 mean observations from each of the 22 horses by 2DE and 3DE methods, by observer 1 (Figure 2). For observer and method associated ICCs observer and method repeatability and reproducibility were determined to be excellent if <10%, good if <25%, moderate if between 26% and 50% and poor if >50%. Agreement between observers in 3DE measurements and for 3DE versus 2DE for obs_1 were also assessed using Bland Altman analyses (using the *BlandAltmanLeh* package[3]).

The relationship between both LAV_{max} and LAV_{min} in the different observers and methods of measurement, were assessed by linear regression. For each horse a single mean 3DE LAV_{max} and 3DE LAV_{min} value for each observer was calculated and then the difference between the observers was calculated, this difference was then regressed against horse bodyweight.

Results

Images selected for further analysis

After grading the image quality, 22 images were used for further analysis, from 21 geldings and 1 mare aged 4-9 years with a bodyweight range from 411-534kg (mean 472kg). The median heart rate was 35bpm (range 27-48bpm). Applying the grading system to all images obtained from all horses resulted in selection of a single image, either obtained by fundamental frequency or harmonic imaging, from each horse to be measured. Twelve selected images were obtained using harmonic imaging, 10 using fundamental frequency. The image volumes per second median for all images was 29.3 (range 15.8-42.1VPS); for harmonic imaging 24.3 (range 15.8-40.5VPS), and for fundamental frequency imaging was 37.8 (range 20.9-42.1VPS). Plots of measurements obtained each day were assessed and evaluation of LAV over the 4 days by both observers revealed no bias from day of measurement (Supplementary Item 3).

3DE volumes - intra-observer variability

The left ventricular quantification package was a feasible technique for left atrial volume measurement. The overall combined fundamental frequency and harmonic imaging median 3DE LAV_{max} for obs_1 was lower compared to obs_2 (596cm^3 versus 852cm^3 respectively, Table 1), and this was also observed in 19 of the 22 horses (Figure 3a). Within observers, there was low intra-observer variability, as assessed by the ICC_{horse} ≥ 76% for both observers (Table 1), and the average standard deviation of the four repeated-measures per horse was 56.6cm^3 for obs_1 and 74.0cm^3 for obs_2, giving *RC's* of 156.8cm^3 and 205cm^3 respectively. There was variation between the observers based on the acquisition method, with the median difference between observers for the data acquired via fundamental frequency lower (obs_1 = 635.5cm^3, obs_2 = 788.5cm^3, difference = 153cm^3, Table 2a, Figure 4a) compared to the harmonic imaging data (obs_1 = 571.0, obs_2 = 893.5, difference = 322.5cm^3, Table 2b, Figure 4b). Intra-observer variability in the acquisition method split data were qualitatively very similar (ICC_{horse} 69 to 85%, Table 2a,b) to the combined data.

The overall median 3DE LAV_{min} measurements for the two observers gave a similar bias as the 3DE LAV_{max}, with median 3DE LAV_{min} for obs_1 lower, than obs_2 (Table 1), and this was observed in 18 of the 22 horses (Figure 3b). Within observers, there was even lower intra-observer variability in the 3DE LAV_{min} data, with ICC_{horse} ≥ 84% for both observers (Table 1), and the average standard deviation of the four repeated-measures per horse was 37.7cm^3 for obs_1 and 38.5cm^3 for obs_2, the *RC* for obs_1 was 104.4cm^3 and for obs_2 was 106.6cm^3. There was less variation between the observers based on the acquisition method, but the median difference between observers for the data acquired via fundamental frequency were still lower (Table 2a, Figure 4c) compared to the harmonic imaging data (Table 2b, Figure 4d). Intra-observer variability in the acquisition method split data were also qualitatively very similar (Table 2a,b) to the combined data.

Left atrial emptying volumes (3DE LA EV) and ejection fractions (3DE LA EF) were also calculated to compare between observers given obs_2 had a bias to measure larger volumes irrespective of acquisition method (Tables 1, 2a,b). Overall median 3DE LA EV and 3DE LA EF were greater in obs_2 compared to obs_1 (Table 1), with similar differences observed in the harmonic imaging data (Table 2b). Whilst 3DE LA EV calculated from fundamental frequency images gave a qualitatively similar result, the median 3DE LA EF of obs_1 was higher than that of *obs*$_2$ (Table 2a). Intra-observer variability was predominately slightly greater for both observers for both 3DE LA EV - with lower ICC_{horse} for combined data (Table 1) and data split into acquired via fundamental frequency (Table 2a) or harmonic imaging (Table 2b)

3DE volumes - inter-observer variability

The $ICC_{inter\text{-}observer}$ for combined 3DE data was 58% for LAV_{max} and 50% for LAV_{min} (Table 1), indicating consistent differences between observers measuring images from the same horse. Bland-Altman analyses revealed that on average there were mean differences of 256cm^3 for 3DE LAV_{max}, and 154cm^3 for 3DE LAV_{min} (Table 3). There was considerable variation between observers for individual horses in both 3DE LAV_{max} and 3DE LAV_{min} (Table 3, Supplementary Item 4). These differences were not related to bodyweight, as there was no statistically significant relationship with the observed differences between observers in both 3DE LAV_{max} ($P = 0.156$, $R^2 = 10.2\%$) and 3DE LAV_{min} ($P = 0.069$, $R^2 = 16.3\%$) (Figure 5).

Fundamental frequency acquired data had a lower 3DE LAV_{max} and 3DE LAV_{min} $ICC_{inter\text{-}observer}$ (≤35%, Table 3a), indicating less consistent differences between observers (Figure 4c), but mean differences were 197cm^3 for 3DE LAV_{max} and 108 for 3DE LAV_{min}, and maximum differences were 764cm^3 and 466cm^3, respectively (data not shown). In contrast, data acquired using harmonic images had a higher 3DE LAV_{max} and 3DE LAV_{min} $ICC_{inter\text{-}observer}$ (Table 3b), indicating more consistent differences between observers (Figure 4d), with mean differences of 304cm^3 and 193cm^3, respectively, but still considerable variation in differences (maximum = 577cm^3 and 392cm^3, respectively, data not shown).

The $ICC_{inter\text{-}observer}$ for combined 3DE LA EV data was much lower (27%) than for 3DE LAV_{max} and 3DE LAV_{min} (Table 1), indicating less consistent differences between the two observers. However, mean differences were 101 cm^3, with wide variation in differences between horses (Table 3, Supplementary Item 4). This relatively lower level of ICC in 3DE LA EV was replicated if the data was split by acquisition method (Table 2a,b), and Bland-Altman analyses showing mean differences of 88-112cm^3 for the two methods (data not shown).

The $ICC_{inter\text{-}observer}$ for 3DE LA EF data whether combined, or divided into fundamental frequency or harmonic images was ~0%, indicating effectively no consistent differences between observers (Tables 1, 2a,b). This is reflected in the mean difference 3DE LA EV of <1% for both combined data (Table 3), and fundamental frequency or harmonic images. However, there was still variation in differences in 3DE LA EF between observers (maximum = 27%, Supplementary Item 4).

2DE volumes – intra-observer variability

The obs_1 overall median 2DE LAV_{max} was higher at 687.5cm^3 compared to the 3DE measurements (596cm^3, Table 1), but this was only observed in 11 of the 22 horses (Figure 6a). As with obs_1 3DE measurements, there was low intra-observer variability for the 2DE (ICC_{horse} = 75%), and the average standard deviation of the four repeated-measures per horse was 63.7cm^3 with a *RC* of 176.5cm^3. However, the differences calculated by obs_1 for 3DE and 2DE were very dependent on the acquisition method, with the median difference between data acquired via fundamental frequency being 1cm^3 (Table 2a), compared to a 132cm^3 for the harmonic imaging data (Table 2b). Intra-observer variability in the acquisition method split data were qualitatively very similar (ICC_{horse} 66-77%, Table 2a,b) to the combined data.

Overall median 2DE LAV_{min} was also higher at 392.5cm^3 compared to the 3DE measurements (373cm^3, Table 1), but again this was not in all horses, being only observed in 10 (Figure 6b). Intra-observer variability was low (ICC_{horse} = 67%), and the average standard deviation of the four repeated-measures per horse was 59.4cm^3, giving a *RC* of 164.5cm^3. Differences between 3DE and 2DE calculated were less dependent on the acquisition method, with median differences in data acquired via fundamental frequency being 11cm^3 (Table 2a), compared to a 29cm^3 for the harmonic imaging data (Table 2b). Intra-observer variability in the acquisition method split data were qualitatively very similar (ICC_{horse} 60-64%, Table 2a,b) to the combined data.

Median 2DE LA EV and 2DE LA EF were also higher compared to the 3DE measurements in combined data (Table 1) and data split by acquisition method (Table 2a,b), However, intra-observer variability was much higher in 2DE LA EV and 2DE LA EF compared to 3DE measurements (ICC_{horse} 46 and 54%, respectively, Table 1), with this again reflected in the data split by acquisition method (Table 2a,b).

3DE vs 2DE volumes – inter-method variability

The $ICC_{inter\text{-}method}$ from comparing 3DE and 2DE was ~14% for LAV_{max} and ~0% for LAV_{min} (Table 1), indicating very little or no consistent differences between 3DE and 2DE as measured by obs_1. LAV_{max} $ICC_{inter\text{-}method}$ dropped to ~0% for fundamental frequency, but rose to 33% for harmonic imaging data (Table 2a,b), whereas for LAV_{min} there was no change in the $ICC_{inter\text{-}method}$ for fundamental frequency data, and a smaller increase in harmonic image data. The smaller $ICC_{inter\text{-}method}$ obtained are reflected in the relatively small mean differences observed from the Bland-Altman analyses (<83cm^3, Table 3). However, there was still considerable variation between methods for individual horses in both LAV_{max} and LAV_{min} (Table 3, Supplementary Item 4).

The $ICC_{inter-method}$ for combined LA EV data was similar to that of LAV_{max} (17%, Table 1), again indicating less consistent differences between the two methods. Mean differences were also lower 64cm^3 (Table 3) but there was still wide variation in differences between horses (Table 3, Supplementary Item 4). These LA EV $ICC_{inter-method}$ values were replicated if the data was split by acquisition method (Table 2a,b). The $ICC_{inter-method}$ obtained for 3DE vs 2DE for LA EF, whether combined, or divided into fundamental frequency or harmonic images was low (≤12%), again indicating very low or no consistent differences between methods (Tables 1, 2a,b). This was reflected in the mean difference LA EV of 5% (Table 3), but there was still variation in differences in LA EF between methods (maximum = 38%, Supplementary Item 4).

Discussion

This study demonstrates that assessment of left atrial volume using software designed for measuring the left ventricle is feasible with good intra-observer repeatability, reflected in the high intra observer ICC, and better intra-observer repeatability compared to 2DE estimates. The average standard deviation from repeated measurements and repeatability coefficients per horse were low, supporting the good intra-observer repeatability. This will be of value when assessing progression of disease in horses, where progressive left atrial dilatation secondary to mitral regurgitation reduces the left atrial emptying fraction and increases left atrial pressure and contractile dysfunction [21]. Left atrial enlargement is also a known risk factor for atrial fibrillation [12, 23, 23] and for recurrence of fibrillation following conversion to sinus rhythm [24]. Reliable measurement of left atrial size also has clear diagnostic and prognostic value in humans [25] and dogs [16].

In humans, measurement of left atrial volume provides a better assessment of left atrial size [14, 15, 21] than linear dimensions which may result in under- or over-estimation of the true diameter of the atrium and may not accurately represent left atrial enlargement in all planes. Cardiac MRI is the gold standard method for assessing left atrial volume but neither this technology, nor cardiac CT, are currently feasible in horses. In this study 3DE LAV was compared to that calculated from a single 2DE image using the single plane method of discs. The AAE/EAE guidelines for estimating LAV from 2DE datasets in humans, recommends acquiring apical views and using either an ellipsoid area-length model or a disc summation algorithm (method of discs) [17] as used in this study. The ellipsoid model is considered less accurate because of reliance and limitations of the linear measurements required to carry out this calculation and the disc summation algorithm is recommended where possible for measuring LAV using 2DE datasets in humans and dogs [16, 17, 26]. The recommended apical approach however, is not possible in horses and 2DE LAV is rarely calculated due to the perceived implications for repeatability of a single parasternal image. Nevertheless, the single plane method of discs was used in this study, calculated from a single 2DE standard right parasternal image orientated to maximise the left atrium. A single plane technique was chosen for 2DE measurements because we wanted to compare 3DE to 2DE with a conventional, standard equine 2DE view and the right parasternal view is currently the most used. There is evidence in human medicine that single plane echocardiography is comparable to biplane 2DE in the evaluation of LAV [27].

Repeat measurement of 2DE images showed more variation in LAV_{min} values than 3DE in the horses in this study. In human cardiology 3DE LAV measurements performed better in terms of correlation, bias, limits of agreement and inter-observer variability [14, 15] and had superior prognostic ability. While reasonable estimates of LAV can be made using 2DE, the inherent geometric assumptions limit the accuracy of such measures [12], particularly in disease where the assumptions may fail. In contrast, 3DE does not make the same assumptions and by encompassing the entire left atrium in the three-dimensional data set, this technique can potentially provide a more accurate assessment of LAV while also reducing variability in different image acquisition planes acquired between different observers. However, despite this theoretical advantage, we cannot comment on variability in 3DE acquisition between observers as only one observer acquired the dataset in this study. The 2DE images from the 3DE dataset are likely to have had lower spatial and temporal resolution compared to images acquired from a 2DE probe which may have contributed to some variation. Nevertheless, the variability noted in this study is consistent with that observed in a study using 2DE datasets for assessing LAV [12].

The left ventricular quantification software (4D Auto LVQ; EchoPAC *v. 202*[2]) used in this study provided a simple, feasible technique to measure LAV. The software and algorithms, designed to measure the left ventricle, were flexible enough to trace the inner side of the left atrial endocardium throughout the cardiac cycle. The main difficulty in using the 4D LVQ package for LAV measurement was that the software expected the maximum volume at end diastole and for this to correspond with the ECG trace. Because atrial volume was being assessed, volumes were largest at end systole and this generated a problem for the algorithms. As a consequence, a technique was adopted to adjust the end diastolic point to the end of atrial diastole, the point at which the largest volume was obtained when the mitral valve was about to open. This adjustment was also required for the minimum LAV. Automated graphs were subjectively assessed for filling and biphasic emptying; normal patterns were observed for all atria included. A pilot study (unpublished data) revealed that if the *Move ED*[2] and *Move ES*[2] buttons were not manually adjusted for each measurement then the automated graph would be inaccurate. During this study, left atrial quantification software for this software package became available, however preliminary evaluation indicated the algorithm was not applicable for measuring the equine left atrium. The reason for this is not clear without more information on the assumptions made by the measurement algorithm.

Of the 40 datasets available, 18 were excluded from the study, because of either a volume rate <15 per second, some of the atrial wall not contained within the data set or a combination of both. At the time of collection, the operators were relatively novice with 3DE acquisition techniques and the large rejection rate in this study emphasises the need to acquire as good quality datasets as possible for analysis. Such good quality acquisition is a significant challenge with 3DE and the operator took several datasets to get the best multi-beat acquisition for analysis, yet there was still considerable rejection of images at subsequent image grading. The LVQ software instructs a minimum of 12 volumes per second for analysis however 15 or greater was used as a lower limit in this study; in most cases, this was significantly higher with a median of 29.3 VPS (range 15.8-42.1VPS). Direct comparison between fundamental and harmonic images in this study was not possible as images were not matched; i.e. images were from different horses. Comparisons on the imaging modality and its effect on measurement repeatability are therefore limited as bias may have been introduced. The horses were not assigned into groups, but categorised purely by initial image quality. At initial acquisition, fundamental frequency and harmonic imaging images were recorded in all horses. Then analysis using the grading system resulted in only single images of either type of imaging modality to be acceptable for measurement. Further study should prospectively compare imaging modalities on 3DE LAV measurement repeatability and reproducibility,

The fit athletic horses in this study were of a single breed from which good quality images could be obtained, which is a limitation of this study. The results of this study may therefore not hold true for datasets from breeds where image quality is poorer. Further study with a larger population of mixed breeds would be required to evaluate the effects of different breeds on image quality and observer repeatability and reproducibility. The horses were imaged unsedated on a yard so that the setting would correspond with a practical clinical situation. The LAV range (Figure 3a,b) was wide which is unsurprising given that this population of athletic horses had different weights, ages [6, 28] and were at different stages of training [29]. It is worth noting also that minimal changes in left atrial diameter (and therefore radius) will have a significant impact on volume measurements, creating a wider range than that observed with linear measurements.

Given the expected range of left atrial diameters of the horses contained in this study, the results for LAV in this study are considered realistic.

In echocardiography, measurements are typically taken over at least three cardiac cycles to provide an average and decrease the variability associated with frame selection. In this study, the multi-beat acquisition, a composite of 4 successive cycles was used. The multi-beat acquisition offers the advantage of enabling the use of harmonic imaging and thus improving image quality for this relatively large volume. Observers measured the same dataset and multi-beat cycle so minimising variation in image acquisition. It may be that greater variation is induced by acquisition between observers although this was not measured in this study. In theory however, 3DE should mitigate against the differences in acquisition between observers, since all image planes are sampled simultaneously. Efforts should be made to acquire the best LAV dataset from a modified parasternal four chamber view, which can be challenging. Of further consideration, the time spent measuring is significantly greater than that of single linear or area measurements; to reproduce a realistic clinical situation therefore, the authors felt that establishing variability from one good measurement was optimal. Single beat acquisition is possible but the image quality and frame rate suffer significantly: both these factors may introduce poorer accuracy and greater observer variability. In the case of arrhythmias such as atrial fibrillation, multi-beat acquisition is not possible. Further studies should therefore assess the repeatability of single beat measurements rather than multi-beat ones.

In humans, when multi-beat images are acquired, subjects are asked to breath hold, to limit the variability attributable to respiratory motion. In a canine study [16], dogs were anaesthetised allowing operator control of breathing. Neither techniques were feasible in this study, nor clinically applicable, and minor artefacts, due to 'stitching' the multi-beat images, have to be accepted in horses. As long as these were subjectively minor, this was deemed acceptable. The lower resting respiratory rate of horses (approximate maximum 33% of the heart rate) helps in this regard, when the multi-beat setting is four. In previous equine studies, allometric scaling has allowed for correction of difference in atrial size associated with bodyweight [6], this was not assessed in this study where the primary aim was to assess repeatability and reproducibility of the technique. The impact of variables such as heart rate, sex, age, intensity of training, image or variability throughout the breathing cycle were not analysed with respect to LAV. Future studies that aim to establish reference ranges for LAV should provide normal values indexed to another morphometric characteristic or intra-cardiac structures (e.g. the aorta).

The intra-class correlation coefficient was elected to be the most desirable way of determining repeatability and reproducibility for intra-, inter-observer and inter-method measurements, reflecting the degree of correlation, repeatability and reproducibility between results [19, 30, 31]. The same observers were repeatedly measuring exactly the same cardiac cycle. To reduce observer bias, measurements were obtained in a random generated order to decrease error that could result from sequential measuring, with observers unaware of previous measurements or horse identification. In this study, repeatability of 3DE measures of LAV by a single person was good, as evidenced by the high ICC for repeated measurements in horses in the same observer. Calculated repeatability coefficients suggest that approximately 200cm^3 change in volume at 3DE LAV_{max} or a 100cm^3 change in volume at 3DE LAV_{min} would suggest a significant enlargement. Lower intra-observer variation was observed for minimum atrial volume compared to maximal atrial volume (Table 1). Reasons for this are unknown but may include better image quality during this stage of the cardiac cycle when the atrium is smaller or a more repeatable frame selection at this cycle stage.

Sources of measurement variation using the 4D LVQ algorithm could have arisen from several sources. Firstly, consistent determination of end systole and end diastole is crucial. Determining the last frame before mitral valve opening relies on some degree of subjectivity, especially when the frame rate is higher. The volume rate per second will have a significant impact on whether the frame chosen is truly the end of diastole/systole. Furthermore, during image analysis, accurate placement of markers at the hinge points of the mitral valve and atrial apex is important. Following creation of the volume boundaries, manual border manipulation may also contribute to variability. The sources of variation noted above are also fundamentally affected by the image quality, determined by volume rate and line density, both of which are compromised when large volumes are required. The effects of image quality were further evaluated by analysing variability between images attained with the fundamental frequency from those obtained using harmonic imaging, which is used to improve image resolution. Intra-observer 3DE variability for both observers was still >75% when only the images obtained with fundamental frequency were analysed. The average volume rate for the fundamental frequency images was higher than that for harmonic imaging as expected, since harmonic imaging compromises temporal resolution. When images obtained by fundamental frequency only were analysed, intra-observer variability decreased (ICC_{horse} increased) for 3DE LAV_{max} for observer one and 3DE LAV_{min} for observer two, and inter-observer and inter-method variability decreased ($ICC_{inter\text{-}observer}$ and $ICC_{inter\text{-}method}$ decreased) suggesting that a higher frame rate decreased measurement variability. In contrast however, variability increased slightly with fundamental frequency for 3DE LAV_{min} for observer one and 3DE LAV_{max} for observer two. Given the design of this study it is difficult to draw firm conclusions about whether fundamental frequency or harmonic imaging is preferred. As a general rule however, obtaining images with the highest resolution and frame rate for a given volume is preferable. Although basic criteria for measurement were set in advance for both observers, it is likely that refined measurement guidelines, being more prescriptive about precise cursor placement, timing of measurement at end systole and diastole, volume border demarcation at the mitral valve and criteria for border manipulation may decrease inter-observer variation. This emphasises the need for different observers to adhere to the same strict measurement guidelines. The study was also limited by the low observer number, further study would be indicated with refined measurement guidelines and more observers of differing expertise to evaluate 3DE and 2DE LAV measurement repeatability, reproducibility and guidelines.

Human studies have determined that 3DE LAV volumes are larger [13, 14, 15, 18, 32] compared to 2DE. In this study, the mean difference for 2DE volumes were greater than 3DE (Table 3) but 2DE did not give consistently higher values than 3DE. Although the accuracy of both techniques could not be established due to the lack of a gold standard method to compare LAV in horses, similarity with comparative studies in humans and dogs, would suggest that 3DE also offers a more accurate measurement of LAV in horses. Despite the low ICC between 2DE and 3DE measurements of LAV, the wide difference limits at individual horse level supports the suggestion that LAV obtained by different methods should not be used interchangeably, a conclusion also reached in a canine study [16]. Fundamental frequency datasets improved inter-method variability to ~0%, suggesting that image quality may be a contributing factor to the variability.

Conclusion

Assessment of LAV by 3DE is feasible and shows good intra-observer repeatability and moderate inter-observer reproducibility. Results using the 2DE method of discs for LAV assessment should not be used interchangeably with that of 3DE. Variability using both techniques for LAV assessment may improve with higher frame rates and refined measurement guidelines but three-dimensional techniques for assessing volume, if available, are likely to be preferable for longitudinal evaluation of left atrial volume due to their better repeatability.

Authors' declarations of interest

No competing interests have been declared.

Ethical animal research

The study protocol was approved by the Veterinary Ethic Research Committee, University of Edinburgh.

Informed consent

Owners gave consent for their animals' inclusion in the study.

Data accessibility statement

The data that support the findings of this study are available from the corresponding author upon reasonable request

Source of funding

The primary author's residency and the study funded by the Horserace Betting Levy Board.

Authorship

F. Worsman had full access to all the data in the study, takes responsibility for integrity of the data and accurate data analysis, contributed to the study execution, data analysis and interpretation, preparation of the manuscript and final approval of the manuscript. Z. Miller contributed to study design, data analysis and interpretation and had final approval of the manuscript. D. Shaw contributed to data analysis and interpretation, preparation and final approval of the manuscript. K. Blissitt contributed to study design, study execution, preparation and final approval of the manuscript. J. Keen contributed to study design, study execution, data interpretation and preparation and final approval of the manuscript.

Manufacturers' addresses

1. 4V-D Active Matrix 4D Volume Phased Array Transducer 1.5-4.0MHz, Vivid E9, GE Healthcare, GE Med Systems Ltd, Chalfont St Giles, Bucks, UK.
2. EchoPac PC version 202, GE Healthcare, Horten, Norway.
3. *lmer, sjPlot* and *BlandAltmanLeh* packages accessed https://www.R-project.org (2020) R version 4.0.0, R Core Team, The R Foundation for Statistical Computing, Vienna, Austria.

Figure Legends

Figure 1: 4 planes of LAV from 3DE dataset (left) showing left atrial volume changes throughout the cardiac cycle. On the right, the time volume curve is enlarged: note the filing to maximal volume at end diastole (*ED*) then the passive and active emptying phases to minimal volume at end systole (*ES*).

Figure 2: Flow chart illustrating study design for Intra-Observer, Inter-Observer and Inter-Method data analysis.

Figure 3: Boxplots with individual values overlaid for obs_1 and obs_2 for a) 3DE LAV_{max} (cm^3) b) 3DE LAV_{min} (cm^3), n = 22 horses. Horizontal black lines indicate the median for that horse from both observers.

Figure 4: Boxplots with individual values overlaid for obs_1 and obs_2 for (a,b) 3DE LAV_{max} (cm^3) and (c,d) LAV_{min} (cm^3) divided into whether fundamental frequency, n = 10 horses (a,c) or harmonic imaging n = 12 horses (b,c) were used to acquire the data.

Figure 5: Relationships between left atrial volumes and bodyweights.

Figure 6: Boxplots with individual values overlaid for obs_1 for 3DE and 2DE measured data for a) LAV_{max} (cm^3) and b) LAV_{min} (cm^3), n = 22 horses.

Supporting information

Supplementary Item 1: 3DE LAV Image Grading System.

Supplementary Item 2: Video: Loop of 3DE LAV_{max}.

Supplementary Item 3: Day-to-day LAV measurement variation.

Supplementary Item 4: Bland Altman plots.

References

1. Borgarelli M., Santilli R.A., Chiavegato D., D'Angolo G., Zanatta R., Mannelli A. and Tarducci A. (2006) Prognostic indicators for dogs with dilated cardiomyopathy. *J Vet Intern Med* **20** 104-10.
2. Borgarelli M., Savarino, P., Crosara S., Santilli R.A., Chiavegato D., Poggi, M. Bellino, C. La Rosa G., Zanatta R., Haggstrom J. and Tarducci A. (2008) Survival characteristics and prognostic variables of dogs with mitral regurgitation attributable to myxomatous valve disease. *J Vet intern Med* **22** 20-8.
3. Rossi A., Cicoira M., Zanolla L., Sandrini R., Golia G., Zardini P. and Enriquez-Sarano M. (2002) Determinants and prognostic value of left atrial volume in patients with dilated cardiomyopathy. *J Am Coll Cardiol* **40** 1425-30.
4. Sargent J., Muzzi R., Mukherjee R., Somarathne S., Schranz K., Stephenson H., Connoly D., Brodbelt D. and Luis Fuentes V. (2015) Echocardiographic predictors of survival in dogs with myxomatous mitral valve disease. *J Vet Cardiol* 2015 **17** 1-12.
5. Wu V.C-C., Takeuchi M., Kuwaki H., Iwataki M., Nagata Y., Otani K., Haruki N., Yoshitani H., Tamura N., Abe H., Negishi K., Lin F. C. and Otsuji Y. (2013) Prognostic value of LA volumes assessed by transthoracic 3D echocardiography. *J Am Coll Cardiol Imaging* 2013 **6** 1025-35.
6. Huesler I. M., Mitchell K. J. and Schwarzwald C. C. (2016) Echocardiographic assessment of left atrial size and function in Warmblood horses: reference intervals, allometric scaling, and agreement of different echocardiographic variables. *J Vet Intern Med* **30** 1241-1252.
7. De Clercq D., van Loon G., Tavernier R., Duchateau L. and Deprez P. (2008) Atrial and ventricular electrical and contractile remodelling and reverse remodelling owing to short-term pacing-induced atrial fibrillation in horses. *J Cardiovascular electrophysiol* **11** 773-784.
8. Schwarzwald C. C. (2019) Equine Echocardiography. *Vet Clin North Am Equine Pract* **35** (1) 43-64.
9. Reef V. B. (1998) Cardiovascular ultrasonography. In Reef V. B., ed. Equine Diagnostic Ultrasound. Philidelphia, PA: WB Saunders 215-272
10. Marr C. M. and Patteson M. (2010) Echocardiography. In: Marr C. M., Bowen M., eds Cardiology of the horse, 2nd ed. Edinburgh: Saunders Elsevier 105-126
11. Vandercasteele T., Cornillie P., van Steenkiste G., Vandevelde K., Gielen I., Vanderperren K. and van Loon G. (2018) Echocardiographic identification of atrial-related

structures and vessels in horses validated by computed tomography of casted hearts. *Equine Vet J* **51** (1) 90-96

12. Schwarzwald C. C., Schober K. E. and Bonagura J. D. (2007) Methods and reliability of echocardiographic assessment of left atrial size and mechanical function in horses. *Am J Vet Res* **68** (7) 735-747
13. Perez de Isla L., Feltes G., Moreno J., Martinez W., Saltijeral A., de Agustin J. A., Gomez de Diego J. J., Maros-Alberca P., Luaces M., Ferreiros J., Garcia Fernandez M. A. and Macaya C. (2014) Quantification of left atrial volumes using three-dimensional wall motion tracking echocardiographic technology: comparison with cardiac magnetic resonance. *Eur Heart J – CV Imaging* **15** 793-799
14. Keller A. M., Gopal A. S. and Kind D. L. (2000) Left and right atrial volume by free hand three-dimensional echocardiography: in vivo validation using magnetic resonance imaging. *Eur J Echocardiogr* **1** (1) 55-65
15. Artang R., Migrino R. Q., Harmann L., Bowers M. and Woods T. D. (2009) Left atrial volume measurement with automated border detection by 3-dimensional echocardiography: comparison with magnetic resonance imaging. *Cardiovasc Ultrasound* **31** 7-16
16. Bouvard J., Thierry F., Culshaw G. J., Schwarz T., Handel I. and Martinez Pereira Y. (2019) Assessment of left atrial volume in dogs: comparisons of two-dimensional and real-time three-dimensional echocardiography with ECG-gated multidetector computed tomography angiography. *J Vet Cardiol* **24** 64-77
17. Lang R. M., Badano L. P., Mor-Avi V., Afilalo J., Armstrong A., Ernande L., Flachskampf F. A., Foster E., Goldstein S. A., Kuznetsova T., Lancellotti P., Muraru D., Picard M. H., Rietzschel E. R., Rudski L., Spencer K. T., Tsang W. and Viogt J. U. (2015) Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. *J Am Soc Echogradiogr* **28** (1) 1-39
18. Muraru D., Badano L. P., Piccoli G., Gianfagna P., Del Mestre L., Ermacora D. and Proclemer A. (2010) Validation of a novel automated border-detection algorithm for rapid and accurate quantitation of left ventricular volumes based on three-dimensional echocardiography. *Eur J Echocardiogr* **11** 359-368
19. Koo K. T. and Li M. Y. (2016) A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropractic Med* **15** 155-163

20. Bland M. An introduction to medical statistics. 3rd Ed. New York: Oxford University Press; 2000 p. 268-93

21. Pagel P. S., Kehl F., Gare M., Hettrick D. A., Kersten J. R. and Warltier D. C. (2003) Mechanical function of the left atrium: new insights based on analysis of pressure-volume relations and Doppler echocardiography. *Anaes* **98** (4) 975-94

22. Schwarzwald, C. C., Hamlin, R. L., Bonagura J. D., Nishijima Y., Meadows C. and Carnes C. A. (2007) Atrial, SA nodal electrophysiology in standing horses: normal findings and electrophysiologic effects of quinidine and diltiazem. *J Vet Intern Med* **21** 166-175

23. Reef V. B., Bonagura J. D., Buhl R., McGurrin M. K. J., Schwarzwald C. C., van Loon G. and Young L. E. (2014) Recommendations for management of equine athletes with cardiovascular abnormalities. *J Vet Intern Med* **28** 749-761

24. De Clercq D., Decloedt A., Sys S. U., Verheyen T., Van Der Vekens N. and Van Loon G. (2014) Atrial fibrillation cycle length and atrial size in horses with and without recurrence of atrial fibrillation after electrical cardioversion. *J Vet Intern Med* **28** (2) 624-629

25. Wu V. C-C. and Takeuchi M. (2017) Three-dimensional echocardiography: current status and real-life applications. *Acta Cardiol Sin* **33** 107-118

26. Wesselowski S., Borgarelli N., Bello N. M. and Abbott J. (2014) Discrepancies in identification of left atrial enlargement using left atrial volume versus left atrial-to-aortic root ration in dogs. *J Vet Intern Med* **28** 1527-1533

27. Vieira-Filho, N.G., Mancuso, F.J.N., Oliveira, W.A.A., Gil, M. A., Fischer C.H., Moises, V.A. and Campos O. (2014) Simplified single plane echocardiography is comparable to conventional biplane two-dimensional echocardiology in the evaluation of left atrial volume: a study validated by three-dimensional echocardiography in 143 adults. *Echocardiogr* **31** (3) 265-272

28. Al-Haidar A., Farnir F., Deleuze S., Sandersen C. F., Leroux A. A., Borde L., Cerri S. and Amory H. (2013) Effect of breed, sex, age and body weight on echocardiographic measurements in the equine species. *Res Vet Sci* **95** 255-260

29. Buhl R., Ersbøll A. K., Eriksen L., Koch J. (2005) Changes over time in echocardiographic measurements in young Standardbred racehorses undergoing training and racing and association with racing performance. *J Am Vet Med Assoc* **226** 1881-1887

30. Popović Z. B. and Thomas J. D. (2017) Assessing observer variability: a user's guide. *Cardiovasc Diagn Ther* **3** 317-324

31. Young L. E. and Scott G. R. (1998) Measurement of cardiac function by transthoracic echocardiography: day to day variability and repeatability in normal Thoroughbred horses. *Equine Vet J* **30** (2) 117-122

32. Badano L. P., Miglioranza M. H., Mihăilă S., Peluso D., Xhaxho J., Marra M. P., Cucchini U., Soriani N., Iliceto S. and Muraru D. (2016) Left Atrial Volumes and Function by Three-Dimensional Echocardiography: Reference Values, Accuracy, Reproducibility, and Comparison With Two-Dimensional Echocardiographic Measurements. Circ Cardiovasc Imaging **9** (7) https://doi.org/10.1161/CIRCIMAGING.115.004229

Table 1: Summary measures (median and range) and Intra-class correlation coefficients (*ICC_{horse}*, *$ICC_{inter\text{-}observer}$*, *$ICC_{inter\text{-}method}$*) of LAV_{max} and LAV_{min} volume, maximum left atrial emptying volume and left atrial ejection fraction for raw data with 4 repeated measurements per horse per observer per method, and average data where the mean value for each set of 4 values is calculated (n = 22 horses).

Data	Volume	Method	Observer				*$ICC_{inter\text{-}observer}$*	obs_1 *$ICC_{inter\text{-}method}$*
			obs_1 Median (range)	*ICC_{horse}*	*obs_2* Median (range)	*ICC_{horse}*		
Raw (4 repeat measurements / horse)	LAV_{max} (cm^3)	3DE	596.0 (349-1029)	77%	852.0 (516-1471)	80%	-	-
		2DE	687.5 (440-1144)	75%	-		-	-
	LAV_{min} (cm^3)	3DE	373.0 (200-721)	84%	533 (314-815)	85%	-	-
		2DE	392.5 (212-776)	67%	-		-	-
	Left Atrial Emptying $Volume_{max}$ (cm^3)	3DE	210.5 (36-542)	70%	311.5 (78-941)	75%	-	-
		2DE	289.0 (4-747)	46%	-		-	-
	Left Atrial Ejection Fraction (%)	3DE	33.2 (5.2-61.0)	76%	36.1 (13.6-64.0)	71%	-	-
		2DE	43.2 (0.5-67.2)	54%	-		-	-
Mean of the 4 repeat measurements / horse	LAV_{max} (cm^3)	3DE	590.4 (405.0-846.0)	-	832.7 (639.8-1214.5)	-	58%	14%
		2DE	689.5 (493.2-1018.8)	-	-		-	
	LAV_{min} (cm^3)	3DE	383.4 (219.2-640.5)	-	528.1 (332.8-777.8)	-	50%	~0%
		2DE	406.0 (239.0-594.2)	-	-		-	
	Left Atrial Emptying $Volume_{max}$	3DE	216.4 (90.8-430.0)	-	318.5 (108.5-742.8)	-	27%	17%
		2DE	285.3 (99.0-447.2)	-	-		-	

	(cm^3)							
	Left Atrial Ejection	3DE	35.3 (14.7-58.3)	-	36.5 (14.9-61.0)	-	~0%	6%
	Fraction (%)	2DE	41.5 (14.6-58.8)	-	-		-	

Table 2: Summary measures (median and range) and Intra-class correlation coefficients (ICC_{horse}, $ICC_{inter\text{-}observer}$, $ICC_{inter\text{-}method}$) of LAV_{max} and min LAV volume, maximum left atrial emptying volume and left atrial ejection fraction for raw data with 4 repeated measurements per horse per observer per method, and average data where the mean value for each set of 4 values is calculated in two different methods of imaging a) fundamental frequency data, n = 10 horses and b) harmonic imaging data, n = 12 horses.

2a. Fundamental Frequency

Data	Volume	Method	Observer *obs₁* Median (range)	*obs₁* ICC_{horse}	*obs₂* Median (range)	*obs₂* ICC_{horse}	$ICC_{inter\text{-}observer}$	obs_1 $ICC_{inter\text{-}method}$
Raw (4 repeat measurements / horse)	LAV_{max} (cm^3)	3DE	635.5 (349-1029)	82%	788.5 (516-1471)	76%	-	-
		2DE	634.5 (440-839)	66%	-		-	-
	LAV_{min} (cm^3)	3DE	350.5 (210-721)	82%	481.5 (314-782)	88%	-	-
		2DE	339.5 (212-524)	64%	-		-	-
	Left Atrial Emptying $Volume_{max}$ (cm^3)	3DE	247.5 (36-433)	61%	297.0 (78-941)	77%	-	-
		2DE	294.5 (58-522)	58%	-		-	-
	Left Atrial Ejection Fraction (%)	3DE	39.4 (5.2-58.0)	60%	36.1 (13.9-64.0)	74%	-	-
		2DE	46.2 (10-67.2)	57%	-		-	-
Mean of the 4 repeat measurements	LAV_{max} (cm^3)	3DE	617.9 (405.0-846.0)	-	758.5 (639.8-1212.5)	-	35%	~0%
		2DE	612.9 (493.2-768.0)	-	-		-	
	LAV_{min}	3DE	345.8 (246.0-640.5)	-	473.4 (332.8-764.0)	-	24%	~0%

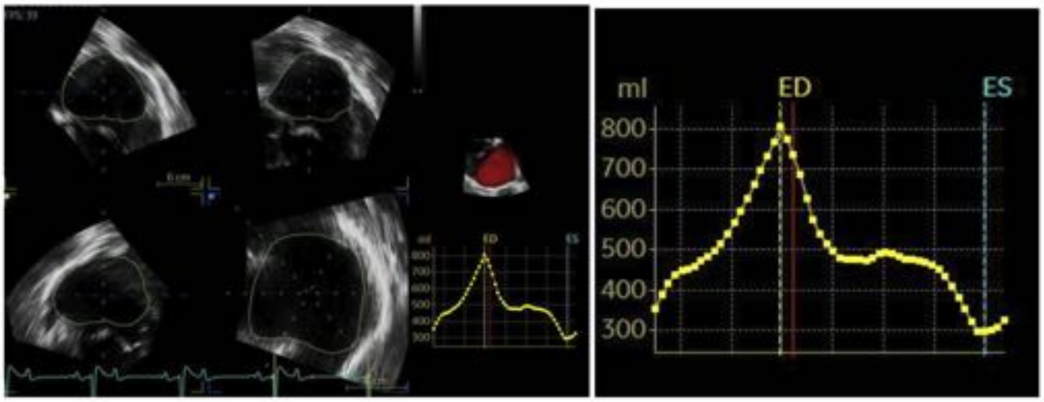
/ horse	(cm^3)	2DE	343.5 (239.0-459.2)	-	-		-	
	Left Atrial Emptying Volume$_{max}$ (cm^3)	3DE	236.5 (110.8-315.2)	-	300.3 (155.5-742.8)	-	12%	12%
		2DE	278.3 (152.8-437.8)	-	-		-	
	Left Atrial Ejection Fraction (%)	3DE	40.8 (14.7-46.8)	-	35.2 (22.6-61.0)	-	~0%	12%
		2DE	45.6 (26.5-58.8)	-	-		-	

2b. Harmonic imaging								
Data	Volume	Method	Observer *obs₁* Median (range)	*obs₁* ICC_{horse}	*obs₂* Median (range)	*obs₂* ICC_{horse}	$ICC_{inter-observer}$	*obs₁* $ICC_{inter-method}$
Raw (4 repeat measurements / horse)	LAV_{max} (cm^3)	3DE	571 (404-888)	69%	893.5 (630-1431)	85%	-	-
		2DE	703 (468-1144)	77%	-		-	-
	LAV_{min} (cm^3)	3DE	386 (200-569)	90%	569.5 (406-815)	80%	-	-
		2DE	415 (220-776)	60%	-		-	-
	Left Atrial Emptying $Volume_{max}$ (cm^3)	3DE	199.5 (64-542)	77%	317.5 (99-648)	75%	-	-
		2DE	279.0 (4-747)	42%	-		-	-
	Left Atrial Ejection Fraction (%)	3DE	31.2 (12.4-61.0)	86%	36.4 (13.6-50.5)	68%	-	-
		2DE	39.8 (0.5-65.3)	49%	-		-	-
Mean of the 4 repeat measurements / horse	LAV_{max} (cm^3)	3DE	590.4 (456.8-754.2)	-	888.7 (651-1214.5)	-	73%	33%
		2DE	702.9 (514.2-1018.8)	-	-		-	
	LAV_{min} (cm^3)	3DE	406 (219.2-540.0)	-	567.9 (420.8-777.8)	-	68%	11%
		2DE	445.1 (263.2-594.2)	-	-		-	
	Left Atrial Emptying $Volume_{max}$ (cm^3)	3DE	173.5 (90.8-430.0)	-	330.8 (108.5-526.8)	-	36%	14%
		2DE	292.4 (99.0-447.2)	-	-		-	
	Left Atrial Ejection	3DE	28.9 (18.5-58.3)	-	37.2 (14.9-43.2)	-	~0%	~0%

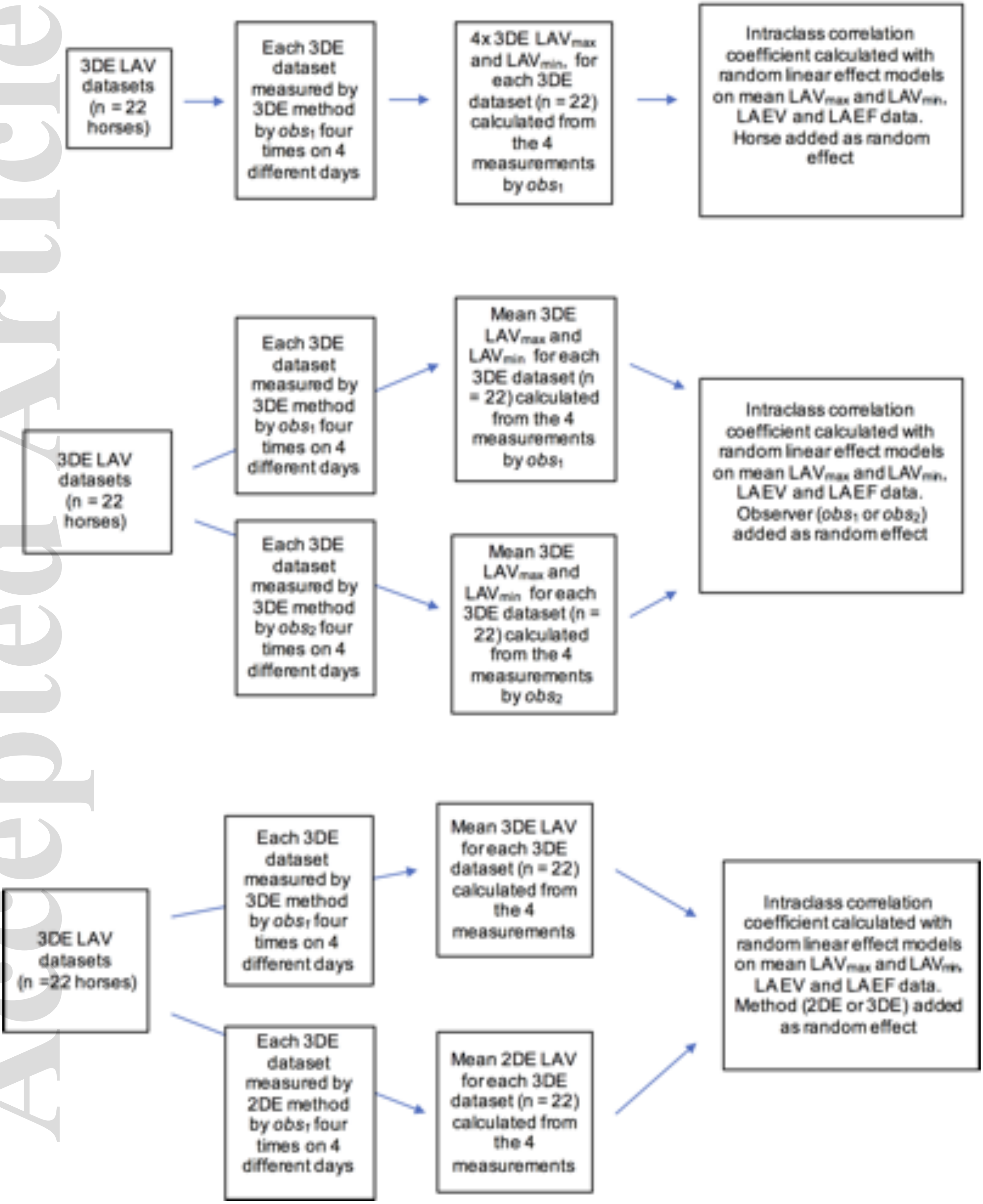
	Fraction (%)	2DE	41.0 (14.6-55.5)	-	-	-	

Table 3: Summary of Bland-Altman analysis of obs_1 3DE measurements in comparison to both obs_2 3DE measurements and obs_1 2DE measurements for LAV_{max} and LAV_{min} volume, maximum Left Atrial Emptying Volume (LA EV_{max}) and Left Atrial Ejection Fraction (LA EF) from the mean value for each set of 4 repeated measurements per horse per observer per method (n = 22 horses).

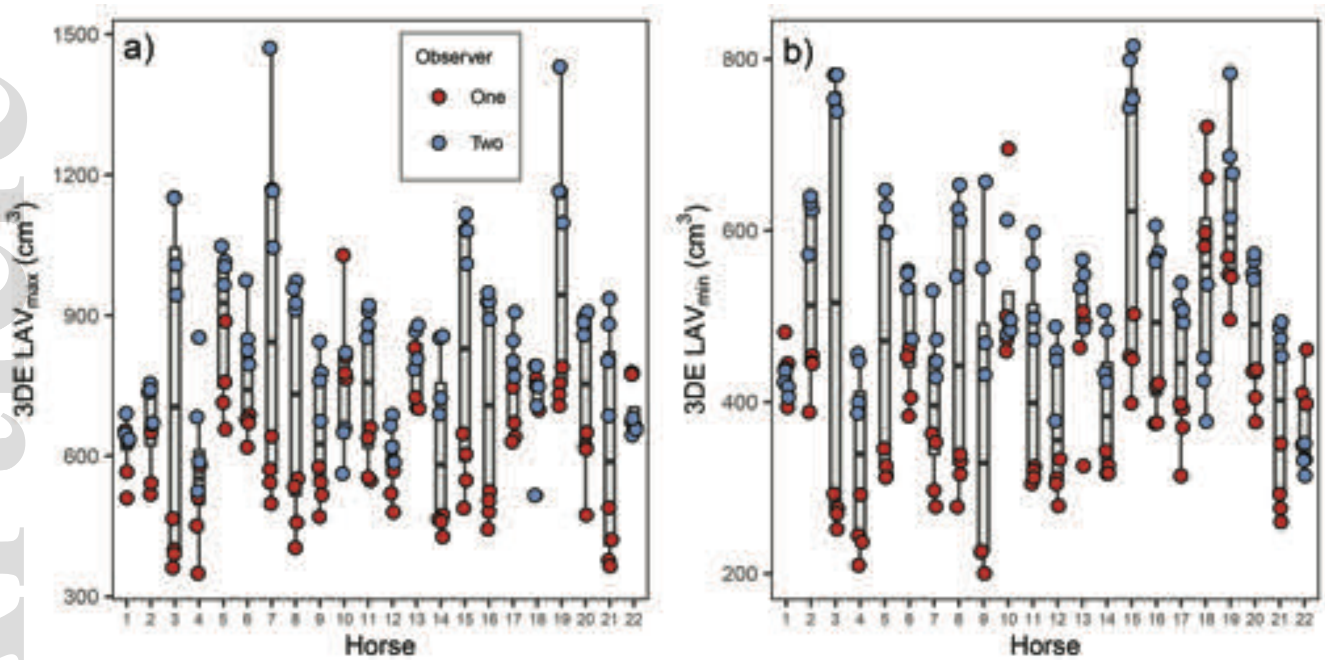
LAV	LAV_{max} (cm^3)	LAV_{min} (cm^3)	LA EV_{max} (cm^3)	LA EF (%)	LAV_{max} (cm^3)	LAV_{min} (cm^3)	LA EV_{max} (cm^3)	LA EF (%)
Method	3DE	3DE	3DE	3DE	3DE vs 2DE	3DE vs 2DE	3DE vs 2DE	3DE vs 2DE
Observer	obs_1 vs obs_2	obs_1 vs obs_2	obs_1 vs obs_2	obs_1 vs obs_2	obs_1	obs_1	obs_1	obs_1
Mean Difference (cm^3)	-255.6	-154.4	-101.1	-0.7	-82.3	-18.3	-64.0	-5.2
Lower Limit (cm^3)	-689.8	-441.7	-393.3	-26.8	-420.2	-299.5	-316.4	-37.7
Upper limit (cm^3)	178.7	132.8	191.0	25.5	255.6	262.9	188.4	27.3



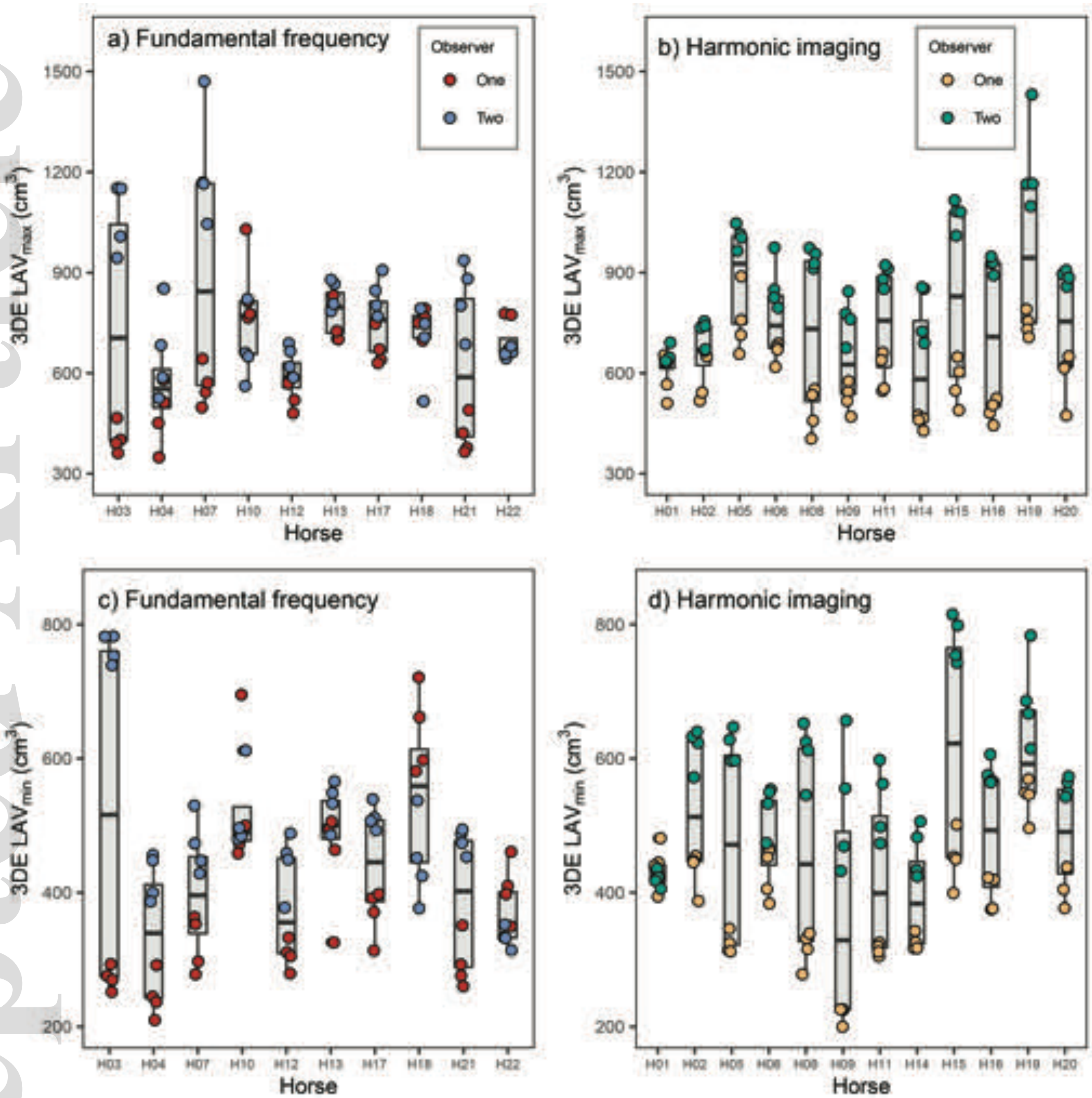
evj_13408_f1.tiff



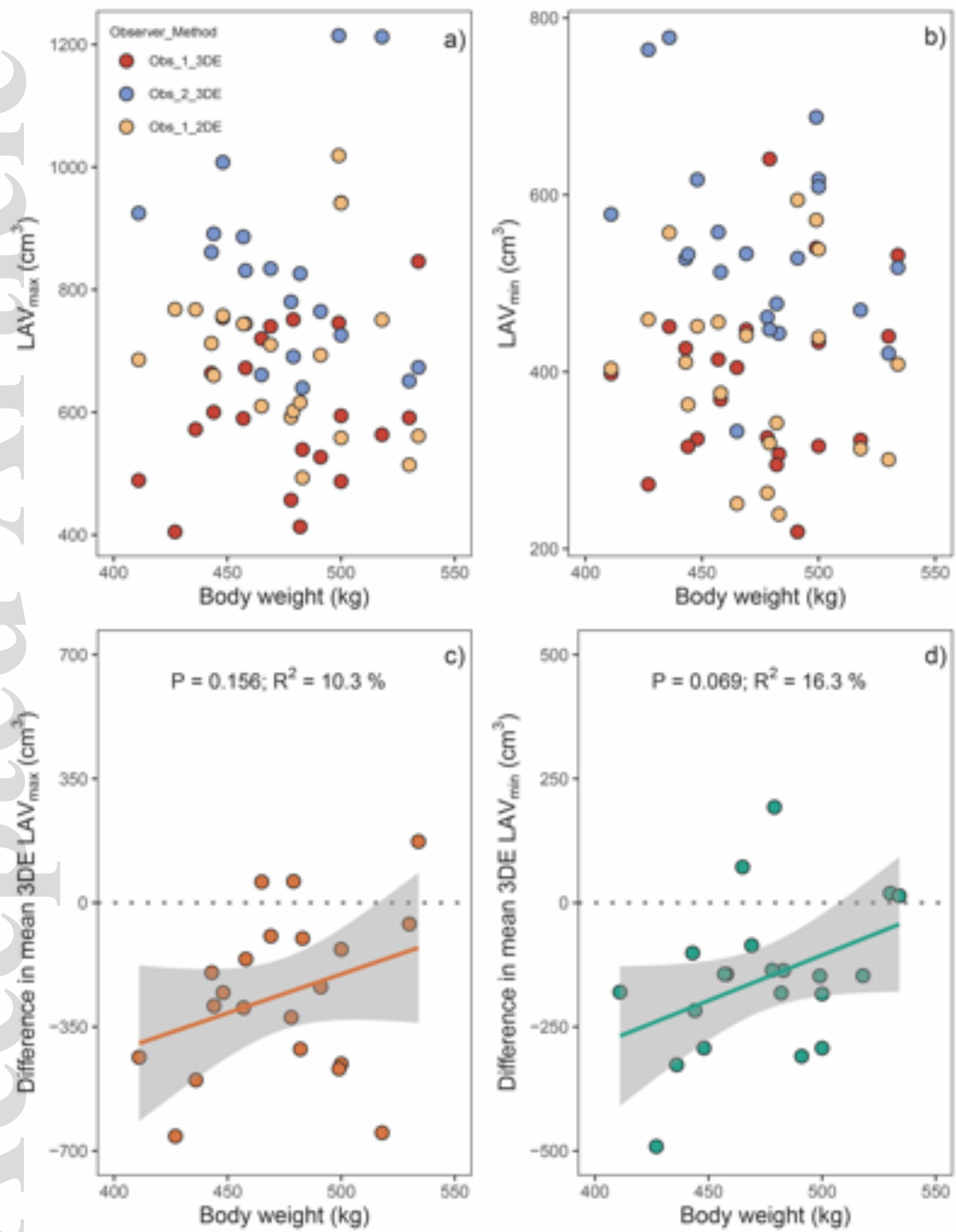
evj_13408_f2.tiff



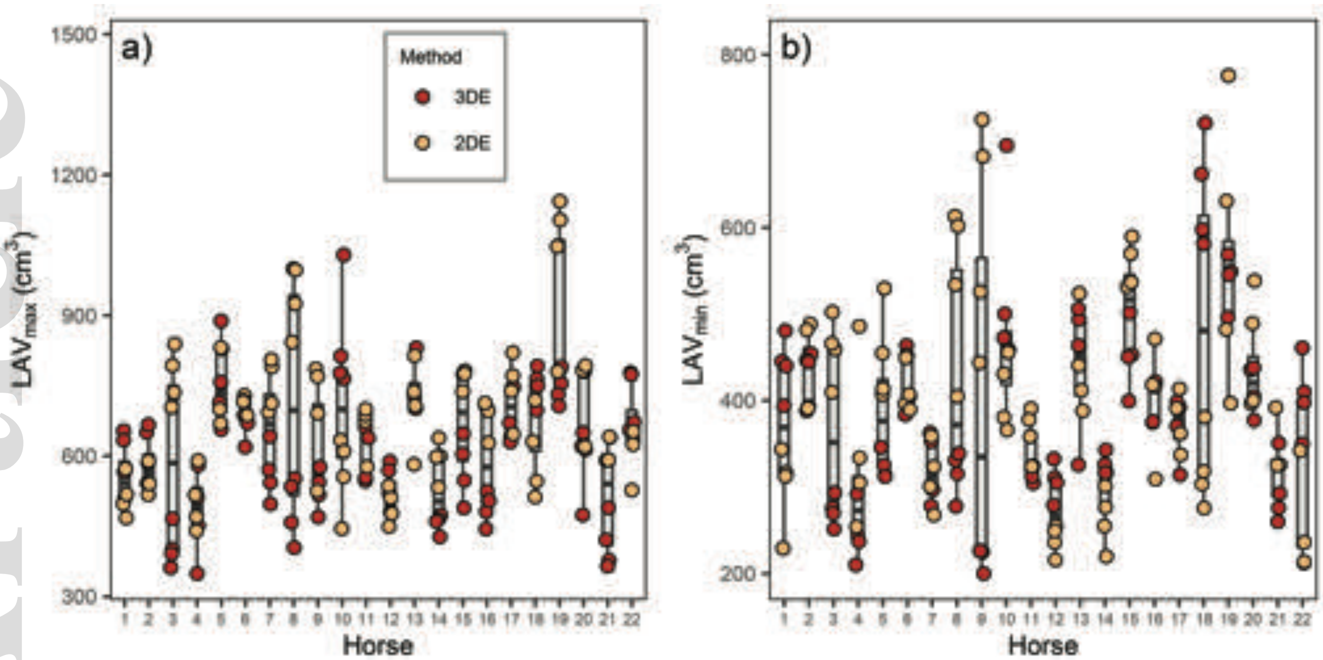
evj_13408_f3.tif



evj_13408_f4.tif



evj_13408_f5.tiff



evj_13408_f6.tif