A comparison of anatomical lateral distal femoral angles obtained with four femoral axis methods in canine femora

J. E. Miles; M. Mortensen; E. L. Svalastoga; T. Eriksen
Department of Veterinary Clinical and Animal Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, Frederiksberg C, Denmark

Keywords
Femur, varus, radiography, reliability, canine

Summary
Objectives: To report the repeatability and reproducibility of four different anatomical proximal femoral axis (a-PFA) methods for measuring anatomical lateral distal femoral angle (a-LDFA), and to compare a-LDFA values produced by each method at three different femoral elevation angles.

Methods: Digital radiographs were obtained of seven dry canine femora at 0°, 12.5° and 25° elevations. Using image analysis software, landmarks defining four different a-PFA and the condylar axis were identified by two independent observers on two separate occasions. Corresponding a-LDFA were calculated for each femur, elevation and a-PFA. Repeatability and reproducibility parameters were calculated and compared statistically, along with the effect of technique and elevation on a-LDFA value.

Results: Interobserver repeatability coefficients were subjectively better for three of the a-PFA methods at 2° compared to the fourth at 3.1°. Median a-LDFA increased significantly (p ≤0.002) with increasing femoral elevation for all a-PFA methods, with a median increase of 3.3°. The median difference in a-LDFA between a-PFA methods yielding the highest and lowest measurements was 2.6° over all three elevations.

Clinical significance: The combined effects of a-PFA choice, femoral elevation and measurement reproducibility may produce typical errors of ± 2.6°, which could have implications for the selection of candidates for corrective osteotomies. Clinicians need to be aware that values obtained with one method and femoral elevation may not be equivalent to values obtained with other methods or elevations.

Introduction
Medial patellar luxation is commonly identified as a cause of lameness in dogs (1). Recent reports have increased the focus on femoral varus as an underlying cause for medial luxation and the use of distal femoral osteotomy as a management approach (2, 3). Criteria for surgical intervention are based on clinical experience and include gross limb deformity and radiographic varus measurements in excess of 10° or 12° (2–4).

Optimal preoperative planning of femoral osteotomy requires a technique that is accurate and precise (reliable). Planning may be performed using computed tomographic reconstructions or using radiographic projections. Radiographically, femoral varus may be assessed as the anatomical lateral distal femoral angle (a-LDFA), calculated as the angle between the anatomical proximal femoral axis (a-PFA) and the condylar axis. Several studies defining a-PFA for calculating femoral varus radiographically have been published (4–7). Three studies have examined the accuracy of radiographic varus measurements; whilst one study found no statistical difference between radiographic measurements and the underlying anatomy, later studies have found poor agreement (4, 5, 8). Acceptable repeatability and reproducibility, as assessed by intra- and interobserver coefficients of variation, have been reported for one a-PFA method in two studies (4, 8).

Radiographic femoral varus is measured from a craniocaudal view of the femur. Correct patient positioning for femoral radiography, avoiding femoral rotation and elevation, is considered important for subsequent determination of femoral varus (4, 5). Positioning the long axis of the femur perpendicular to the beam may not always be possible due to coxofemoral joint disease or other limitations (7). Increasing femoral elevation has been shown to significantly increase measured a-LDFA in one study, but not in another using a different a-PFA method (4, 7). It remains unclear...
how different a-PFA methods will affect a-LDFA measurements under varying degrees of femoral elevation that might be expected in clinical practice.

Each of the published studies describing varus measurement uses different landmarks to define the a-PFA method used for calculating the a-LDFA and the necessary varus correction (Figure 1) (4–6). Variations in the defined a-PFA method could, if substantial enough, result in misapplication of the proposed cut-off values for surgical intervention. In addition, the proximal and distal landmarks for these axes are not widely separated, particularly in shorter femora, and this might negatively impact measurement precision. Based on perceived shortcomings of the published anatomical axes, an additional a-PFA, broadly based on symmetric axis principles described for femoral neck inclination measurement, was introduced for this study (9). Based on the wide separation of its landmarks and presumed simplicity of landmark identification, improved measurement precision was expected compared to the previously defined axes.

The aims of this study were to compare radiographic measurement repeatability and reproducibility for four a-PFA methods, and to investigate a-LDFA agreement between and within these a-PFA methods with increasing femoral elevation.

Materials and methods

Seven canine right femora were obtained from the Department of Veterinary Clinical and Animal Sciences at the University of Copenhagen (Anatomy Section); the femora originated from skeletally mature dogs and none showed evidence of prior surgery or fracture. A minimum sample size of five femora had been estimated as sufficient to yield statistically significant differences based on power calculations using previously published data, assuming high correlations between a-LDFA measurements at different elevations (7).

All femora were digitally radiographed at three elevations (0°, 12.5°, 25°). These elevations were defined by a line connecting the lesser trochanter proximally and the supracondylar tuberosity distally, as previously described (4, 7). Elevation was achieved by use of firm foam wedges custom cut to the required angles, with additional support under the proximal femur as required, and was checked using a gonio-meter. The caudal aspects of the femoral condyles rested on either the radiographic plate or on the firm foam wedge to prevent rotation about the long axis of the femur, ensuring mediolateral superposition of the caudal aspects as previously recommended for a true craniocaudal view of the femur (5). Radiographs were centred on the mid-diaphysis. Images were stored on a picture archiving and communication system as digital imaging and communications in medicine (DICOM) files.

Using both commercial and freely available software, all DICOM files were anonymized and randomized for two readings on separate occasions one week apart by two independent observers. The two observers differed in their experience levels (new graduate and experienced clinician). Readings were performed using freely available imaging software using the multipoint measurement tool. Based on the descriptions in Figure 1, eleven points were identified on each femur for determination of the four a-PFA and the distal condylar axis. The landmark halfway along the diaphysis was shared by two of the methods. Point location was aided by custom transparencies, which contained concentric circles and guides for division of femoral length into halves and thirds, and by the linear measurement tool within the program.

The previously defined a-PFA measured here were coded according to the first author of the publication in which they were described: TO, SW, and DU (4, 5, 7). The fourth a-PFA was coded as SY (Symax).

Coordinate data were exported to a spreadsheet for calculation of the a-LDFA. Where the femoral axis was defined by three points (DU, SW), a least squares approach was used to find the best fit axis.

Femoral length was measured in a single reading from the radiographs at 0° elevation, using the proximal aspect of the greater trochanter and the proximal aspect of the fossa of the long digital extensor muscle on the lateral condyle as landmarks (10).

Statistical methods

Randomisation and blinding were broken, and data analysed using statistical software.

Within-subject standard deviations (WSSD) and repeatability coefficients (RC), and two-way random single measures intra-class correlation coefficients for absolute agreement (ICC2,1a) were used to assess measurement repeatability and reproducibility for each a-PFA method (11). Data were assessed as homoscedastic using Koenker’s test prior to calculation of WSSD. Friedman’s analysis of variance (ANOVA) with post-hoc pairwise comparisons and correction of the significance levels was used to compare the underlying variances of the intra- and interobserver WSSD values. Statistical significance was set at $p = 0.05$. The ICC2,1a values were interpreted according to published criteria (12).

Comparisons of a-LDFA between a-PFA methods for each elevation and between elevations for each a-PFA method were performed using Friedman’s ANOVA with post-hoc pairwise comparisons and correction of the significance levels to compensate for multiple comparisons.

For estimation of overall typical error for measurement (due to the combined effects of a-PFA method, femoral elevation and observer differences), a two-way repeated measures ANOVA was performed and error variances for a-PFA method, femoral elevation and their interaction identified for use with the interobserver variance calculated during derivation of WSSD.
Results

Median femoral length was 19.4 cm (interquartile range [IQR]: 18.4 cm – 20.5 cm). Median values for the a-LDFA varied with both a-PFA method and elevation as shown in Figure 2.

The underlying variances used to calculate WSSD were not significantly different for all elevations and for all a-PFA methods for each observer (p >0.05). Differences in the underlying variances were also not significantly different for the interobserver comparisons except for DU between elevations 0° and 25° and for SY between elevations 12.5° and 25° (corrected p = 0.048 and p = 0.023, respectively). In both cases, estimated WSSD values exhibited substantial overlapping of their confidence intervals. Therefore, pooled WSSD, RC and ICC \(_{2,1A}\) values for intra-observer repeatability and interobserver reproducibility for each a-PFA method over all three elevations are summarized in Table 1. All ICC \(_{2,1A}\) values exceeded 0.81, indicating substantial agreement (12).

In general, the differences in repeatability and reproducibility measures between a-PFA methods were small. Despite a trend for higher WSSD values for method DU, the pooled underlying variances between methods were not significantly different for either observer (p = 0.071 and p = 0.276) or the interobserver results (p = 0.31).

Measured a-LDFA increased with increasing femoral elevation for all a-PFA methods (p ≤0.002). The median increase across all methods from 0° to 25° was 3.3° (IQR: 2.5°–4.3°; estimated 95% confidence interval for the median: 2.6°; 4.0°). Pairwise comparisons, following correction for multiple comparisons, showed that only the increase from 0° to 25° elevation was significant (corrected p ≤0.002).

The choice of a-PFA method affected measured a-LDFA at all elevations (p <0.001). Measurements using methods DU and SY produced lower a-LDFA values.

Figure 1 Four methods for defining the anatomic proximal femoral axis (a-PFA). The four methods used in this study for determining the anatomic lateral distal femoral angle (a-LDFA) are shown. A) Method TO: a-PFA defined by diaphyseal midpoints at one-third and one-half of the proximo-distal femoral length (6). B) Method SW: a-PFA defined by circles centred at the femoral isthmus (distally), an isthmus diameter distal to the lesser trochanter (proximally) and halfway between these two (4). C) Method DU: a-PFA defined by diaphyseal midpoints 1 cm apart, starting immediately distal to the lesser trochanter (5). D) Method SY: a-PFA defined by circles centred at the proximal metaphysis and the isthmus. In all cases, the condylar axis is defined by the distal most aspects of the articular surfaces of the femoral condyles. The a-LDFA is shown as an arc for each method.
J. E. Miles et al.: A comparison of four femoral axis definitions

Intra- and interobserver reliability measures for the four anatomical proximal femoral axis methods (a-PFA). The axes are those illustrated in Figure 1. Within-subject standard deviations (WSSD), repeatability coefficients (RC), and intra-class correlation coefficients (ICC$_{2,1A}$) are shown for observers 1 and 2, and for the interobserver comparison, along with 95% confidence intervals. Pooled values for each method (ignoring femoral elevation) are shown. Values have been rounded.

Figure 1. Variation of the anatomical lateral distal femoral angle (a-LDFA) with both the elevation and axis method. Median values of a-LDFA for the four anatomical proximal femoral axis (a-PFA) methods as defined in Figure 1 are shown for the three femoral elevation angles. Values for a-LDFA at 25° elevation were significantly greater than at 0° for all a-PFA methods, and values for methods DU and SY were significantly lower than for method TO at all elevations. Greenhouse-Geisser corrected mean squares were used as the basis for the typical error calculation because epsilon values were <0.75 despite Mauchley’s test not indicating any deviation from sphericity (p >0.05). Variances of 4.81 (elevation), 1.03 (method) and 0.14 (interaction), along with the mean interobserver variance (0.67) yielded a typical error for measurement of ± 2.6°.

Table 1

<table>
<thead>
<tr>
<th>Observer</th>
<th>a-PFA method</th>
<th>TO</th>
<th>SW</th>
<th>DU</th>
<th>SY</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>WSSD (°)</td>
<td>0.6 (0.5–0.8)</td>
<td>0.6 (0.4–0.8)</td>
<td>1.1 (0.8–1.3)</td>
<td>0.7 (0.5–0.8)</td>
</tr>
<tr>
<td></td>
<td>RC (°)</td>
<td>1.8 (1.3–2.3)</td>
<td>1.7 (1.2–2.1)</td>
<td>3.0 (2.2–3.8)</td>
<td>1.8 (1.3–2.3)</td>
</tr>
<tr>
<td></td>
<td>ICC</td>
<td>0.96 (0.90–0.98)</td>
<td>0.97 (0.92–0.99)</td>
<td>0.94 (0.85–0.97)</td>
<td>0.96 (0.89–0.99)</td>
</tr>
<tr>
<td>2</td>
<td>WSSD (°)</td>
<td>0.3 (0.2–0.4)</td>
<td>0.3 (0.2–0.4)</td>
<td>0.4 (0.3–0.5)</td>
<td>0.4 (0.3–0.5)</td>
</tr>
<tr>
<td></td>
<td>RC (°)</td>
<td>0.9 (0.7–1.2)</td>
<td>0.8 (0.6–1.0)</td>
<td>1.2 (0.9–1.5)</td>
<td>1.1 (0.8–1.4)</td>
</tr>
<tr>
<td></td>
<td>ICC</td>
<td>0.99 (0.98–1.0)</td>
<td>0.99 (0.99–1.0)</td>
<td>0.99 (0.97–1.0)</td>
<td>0.99 (0.97–1.0)</td>
</tr>
<tr>
<td>Inter-observer</td>
<td>WSSD (°)</td>
<td>0.7 (0.5–0.9)</td>
<td>0.7 (0.5–0.9)</td>
<td>1.1 (0.8–1.4)</td>
<td>0.7 (0.5–0.9)</td>
</tr>
<tr>
<td></td>
<td>RC (°)</td>
<td>2.0 (1.5–2.6)</td>
<td>2.0 (1.5–2.6)</td>
<td>3.1 (2.2–3.9)</td>
<td>2.0 (1.4–2.5)</td>
</tr>
<tr>
<td></td>
<td>ICC</td>
<td>0.98 (0.94–0.99)</td>
<td>0.98 (0.94–0.99)</td>
<td>0.97 (0.93–0.99)</td>
<td>0.99 (0.96–0.99)</td>
</tr>
</tbody>
</table>

Discussion

Key findings in this study were the relatively homogeneous reliabilities of the four a-PFA methods for measuring the a-LDFA, and the significant differences in the angles obtained both between a-PFA methods and between elevations. No clear advantage in repeatability or reproducibility could be demonstrated for the new a-PFA method (SY) with wider landmark separation. Since the condylar axis was constant between a-PFA for each elevation, the reliabilities and differences reported here represent true comparisons between the a-PFA methods.

Previous studies of repeatability and reproducibility of radiographic varus measurement have used coefficients of variation (4, 8). Approximate back conversion to WSSD can be performed (since CV=WSSD / mean angle). Given published intra-observer coefficients of variation of 11%, 12% and 16% and interobserver coefficient of variation of 16%, WSSD values similar to ours at 0.6°, 0.7°, 0.9° and 0.9° can be calculated (4). We have avoided the use of this measure for two reasons. Firstly, the value of the coefficient of variation is highly dependent on the denominator of the fraction; if the a-LDFA values were corrected to varus angles by subtraction of 90°, some of our values cross or approach zero, rendering calculation of the coefficient of variation unreliable or impossible. Secondly, the value obtained is a percentage, which is often not intuitively interpreted and which is only valid for the range of values from which the coefficient was calculated. Similar WSSD values could result in widely varying coefficient of variation estimates for mean angles of 5° and 10°. For homoscedastic data (for which error magnitude is unrelated to measurement magnitude), we believe that parameters such as WSSD or RC are to be preferred, since the values obtained are directly relevant to measurement repeatability or reproducibility and are given in the same units as the measurement. The only other study reporting repeatability and reproducibility of a-LDFA measurements is a study of CT reconstructions, which used ICC values (13). Reported values were 0.94–0.98 (intra-observer) and 0.98 (inter-
observer), which are very similar to those found here. It is important to bear in mind that high variability in the underlying data can unfortunately result in masking of quite large variations between measurements when using any correlation-based technique, which can make direct comparisons of ICC values difficult (14).

The RC defines an interval within which 95% of any paired measurements should fall, disregarding directionality. Consequently it is possible to define a range around any one measurement of ± RC within which one can expect the majority of subsequent measurements to fall. Our estimates indicate a range of 4.0°-6.2°, depending on the a-PFA method employed, suggesting that rigid application of cut-off values may be inappropriate. Refinement of the confidence intervals for these estimates will require larger sample sizes: reducing the confidence interval by half would require a combined sample size of around 40, or an increase in repeat measurements to five.

Of the a-PFA methods studied here, two use only two measurement points, and two use a best fit approach using three measurement points. While a best fit approach should potentially minimize errors, we are not aware of any studies confirming this in assessment of radiographic landmarks, and our repeatability and reproducibility measures do not show a clear pattern that could resolve this question. It could be argued that using only two landmarks is simpler. From a geometric error perspective there is a clear advantage in separating the landmarks as far as possible (as for SY), but again this potential advantage is not supported by our data. Although we expected the wide separation of the landmarks to be an advantage compared to the other a-PFA methods, intra-observer RC values for SY were not significantly different to the other three methods.

All a-PFA methods studied produced similar variations in a-LDFA with changing femoral elevation. The elevations in this study were based on the optimal 0° femoral elevation for varus measurement and elevations which might be expected in clinical practice (4-7). Apart from method SW, a-LDFA increased non-linearly with increasing femoral elevation. Whether this represents a clinical advantage is unclear. In terms of a-PFA agreement, additional differences of up to 3.5° could be observed between the different methods depending on femoral elevation; similar effects of elevation on a-LDFA measurement have been reported previously (7). The combination of varying femoral elevation and choice of a-PFA could result in clinically significant differences in a-LDFA measurements. To minimize this problem, clinicians should strive for optimal femoral positioning; in addition, when proposing clinical intervention cut-offs it would be wise to specify the a-PFA used to establish these.

This study has some limitations. The sample size was small, and although sufficient statistical power was achieved for comparing the effect of a-PFA method and elevation on a-LDFA, the sample size has a negative impact on the estimation of 95% confidence limits for the repeatability and reproducibility measurements. Although we identified significant differences in the underlying variances with varying elevation for two methods (DU, SY) we chose to present pooled values based on an assessment of the confidence limits for the WSSD values for each elevation. The differences we identified may represent type I error, or may represent a true variation in repeatability or reproducibility as femoral elevation changes; it is however unclear why this should affect these two methods and not the others. As such, the values presented here should be treated as initial estimates, and confirmation in additional, larger studies is recommended. The positioning protocol employed here controlled for rotation of the femur about its long axis (as have previous studies) and thus our results may present an optimistic view of measurement repeatability and reproducibility in clinical practice, since complete absence of femoral rotation is rarely ensured during radiography. Additionally, the effect of error in defining the condylar axis was excluded by virtue of the study design. The effects of these additional factors on measurement repeatability and reproducibility were not explored here. The increasing availability of CT, particularly in centres likely to be performing corrective femoral osteotomies, and the ability of this modality to correct for elevation and rotation via multiplanar reconstruction, may render these concerns irrelevant. It is not possible from this study to say which a-PFA method is most valid, that is, most accurately represents the true a-PFA. We concur with others in stating that no such gold standard has yet been defined (13). However, the key parameter for surgical planning is precision, since the clinician is primarily interested in reproducibly identifying the difference between normal and abnormal – in this context, accuracy becomes less important (13).

Conclusions

The combined effects of a-PFA choice, femoral elevation and measurement reproducibility may produce typical errors of ± 2.6° which could have implications for the selection of candidates for corrective osteotomies. Clinicians need to be aware that values obtained with one method and femoral elevation may not be directly comparable with values obtained with alternative methods and different elevations.

Acknowledgements

Preliminary results from this study were presented as a clinical research abstract and poster at the British Small Animal Veterinary Association Congress, 3rd-7th April 2013.

Conflict of interest

None declared.

References


© Schattauer 2015

Vet Comp Orthop Traumatol 3/2015