An assessment of the agreement between the New Zealand Veterinary Association Hip Dysplasia Scoring System and the PennHIP distraction index in German Shepherd dogs

AJ Worth, RA Laven & VH Erceg

To cite this article: AJ Worth, RA Laven & VH Erceg (2009) An assessment of the agreement between the New Zealand Veterinary Association Hip Dysplasia Scoring System and the PennHIP distraction index in German Shepherd dogs, New Zealand Veterinary Journal, 57:6, 338-345, DOI: 10.1080/00480169.2009.64720

To link to this article: http://dx.doi.org/10.1080/00480169.2009.64720

Published online: 16 Feb 2011.

Article views: 128

Citing articles: 5 View citing articles
An assessment of the agreement between the New Zealand Veterinary Association Hip Dysplasia Scoring System and the PennHIP Distraction Index in German Shepherd dogs

AJ Worth *, RA Laven † and VH Erceg ‡

Abstract

AIM: To determine the level of agreement between the New Zealand Veterinary Association (NZVA) Hip Dysplasia Scoring System and the University of Pennsylvania Hip Improvement Program (PennHIP) Distraction Index in German Shepherd dogs, and whether using the NZVA subtotal score or its components affected the level of agreement.

METHODS: A prospective study was performed using 47 German Shepherd police dogs undergoing breeding evaluation. All dogs were scored using the NZVA system and the PennHIP index. The relationships between the individual hip-distraction index scores and the scores from the NZVA system, i.e. the total score, the subtotal score, and the scores for the categories making up the subtotal score, were analysed using correlation, followed by univariate ANOVA for the subtotal categories alone. The scores from the NZVA system and the distraction index were then dichotomised into either low or high risk of canine hip dysplasia (CHD). A sign test was then used to determine whether three NZVA thresholds identified the same proportion of ‘at-risk’ dogs as the distraction-index threshold. Where this was the case, the Kappa value was calculated to identify the degree of agreement between the NZVA measures and the distraction index.

RESULTS: The left-hip distraction index was significantly correlated to both left-subtotal and left-total NZVA score, however for right-hip scores there was no such correlation. The individual categories of the subtotal NZVA score were not significantly associated with the distraction index except for the subluxation score of the left hip.

The proportion of dogs identified as being ‘at risk’ for CHD identified using a distraction-index threshold of 0.3 was similar to that identified by an NZVA total or subtotal score of >2 (44/47 cf. 45/47, respectively). However, none of the dogs identified as low risk using the distraction index was identified as low risk by either of the NZVA scores. This poor agreement (Kappa value <0) was not improved by using a threshold of >1 in any of the categories used to calculate the NZVA score. The agreement between the two scores was improved by using different thresholds; distraction index >0.5 and an NZVA total score of >9. These thresholds identified a much lower proportion of ‘at-risk’ dogs (5/47 and 10/47, respectively) than the standard thresholds. The Kappa values using the thresholds of 0.5 for the distraction-index and ≥9 for the NZVA scores were moderate for subtotal (0.55; 95% CI=0.16–0.94) and low for total (0.31; 95% CI=–0.02 to 0.63) score.

CONCLUSIONS: The low level of agreement between the NZVA total scores and the PennHIP distraction index, particularly when the standard interpretations were used, is of concern as assessing dogs by each method gave disparate results. There was evidence from this study that using the subtotal score moderately improved the level of agreement with distraction index, but only at the higher thresholds.

CLINICAL RELEVANCE: The low level of agreement between NZVA and PennHIP results in the same dog precludes them being used interchangeably to guide breeding decisions. The higher heritability of distraction-index measurement in previous studies suggests that it is a better selection tool for breeding dogs when CHD is present within a population. The advantage of a hip-extended ventrodorsal view is its low cost and widespread availability but comparisons between individuals may not be accurate due to the poor sensitivity and the presence of false negatives.

KEY WORDS: Canine hip dysplasia, PennHIP, distraction index, NZVA, hip scores, laxity

Introduction

Canine hip dysplasia (CHD), a developmental trait primarily affecting medium-sized and large-breed dogs, is characterised by instability of the hip joint, leading to degenerative joint disease (DJD), and is usually bilateral (Todhunter and Lust 2003). It may be detected radiographically as subluxation of the femoral head in affected young dogs, or as DJD of the hip joints in older dogs (Todhunter and Lust 2003). CHD is recognised as one of the most common orthopaedic diseases affecting companion animals, and has a polygenetic mode of inheritance (Cook et al. 1996).

Radiographic methods have been developed in an attempt to allow dog breeders to select less phenotypically affected stock for
breeding (Lawson 1963; Smith et al. 1990; Fluckiger et al. 1999; Lust et al. 2001). The hip-extended ventrodorsal view recommended by the Orthopedic Foundation for Animals (OFA) is the radiographic projection most widely used for the determination of CHD status (www.ofa.org). The OFA is a not-for-profit organization dedicated to the reduction of orthopaedic and genetic disease in dogs and cats, and has been established for >40 years. The OFA recommended view of the coxofemoral joints allows consistent determination of the level of osteophytic development, which is regarded as pathognomonic for CHD. The extent of incongruity and subluxation of the coxofemoral joint is assessed subjectively. The phenotypic evaluation by the OFA falls into seven different categories, viz normal (excellent, good, fair), borderline, and dysplastic (mild, moderate, severe).

The Willis CHD scoring system was developed by Dr Malcolm Willis in the United Kingdom, using the hip-extended ventrodorsal view but applying an objective grade to the indicators of subluxation, incongruity and degeneration. There are eight criteria, each scored from 0 to 6, and one criterion scored 0 to 5, totalling from 0 to 53 per hip. Equivalence between the OFA grades and the Willis system is reported as: excellent = 0–4 (not >3/hip), good = 5–10 (not >6/hip), fair = 11–18, borderline = 19–25, mild = 26–35, moderate = 36–50, severe = 51–106. The Willis system was adopted by the NZVA in 1967 as a screening method for the detection of CHD. Statistical analysis of the data held by the NZVA has not been reported, however the mean CHD score for German Shepherd dogs (by year of birth) has lowered from 14.4 to 9.8 in the period 1994 to 2006, although the median only changed from 8 to 7 in the same period (AJ Worth, unpublished obs.). In other countries using similar radiographic-scoring methods, there has been only slow progress in reducing the prevalence of CHD (Corley and Hogan 1985). The NZVA scoring scheme itself is an advisory tool, not a control scheme for genetic disease. Non-enforced selection thresholds, selective use by breeders, and failure to submit radiographs from obviously phenotypically abnormal individuals (pre-screening) all bias the reported incidence and average score of populations measured by such non-mandatory schemes (Paster et al. 2005).

Henricson et al. (1966) described a link between early joint laxity and later development of CHD. It is thus important that any CHD screening tool is able to detect hip laxity, the forerunner of DJD. Previous studies have shown a correlation between Norberg’s angle and the presence of DJD (Smith et al. 1993, 1995). The Willis system includes grades for Norberg’s angle, subluxation (femoral head coverage), and cranial acetabular incongruity as measures of laxity. However, when the coxofemoral joints are positioned in extension, as is done for the hip-extended NZVA/OFA view, the joint capsule and ligament of the head of the femur are both twisted, tightening the tensile elements of the joint capsule, which can mask hip laxity (Smith et al. 1990). Laxity is defined as active when an animal is weight-bearing and the muscles supporting the hips are active, and passive when an animal is chemically immobilised and the maximum laxity measured whilst distraction of the hip is artificially maintained. The work by Smith et al. (1990, 1995) at the University of Pennsylvania has shown that passive laxity is better correlated to the development of DJD and clinical disease than the OFA score. The University of Pennsylvania Hip Improvement Program (PennHIP) uses an empirical method that measures passive laxity, using distraction radiography to generate a distraction index (Smith et al. 1990), and became commercially available through training courses in 1993 (www.pennhip.org). A hip-extended radiograph is submitted to the PennHIP Analysis Center, together with compression/distraction views. The heritability of CHD is reported to be between 0.22 and 0.43 using the hip-extended method (Leighton et al. 1977; Hedhammer et al. 1979), and up to 0.83 using the PennHIP method (Cook et al. 1996; Bliss et al. 2002; Ginja et al. 2008b). The PennHIP network consists of trained veterinarians in 36 countries, including Australia and New Zealand. No national veterinary medical association has yet adopted PennHIP as the only hip-screening system.

In an effort to increase the emphasis on hip laxity, the NZVA CHD system was recently modified by the reporting of a subtotal or ‘subluxation score’ (Burbidge 2003). Rather than only reporting the total score (out of a possible 106), the three criteria deemed to be most indicative of joint laxity, i.e. Norberg’s angle, subluxation score, and incongruity between the femoral head and cranial acetabular edge, are now subtotalled out of a maximum of 36. The NZVA CHD submission form reads thus: “Whilst the ideal score (subtotal) is 0, a 2 or less is acceptable”. This subtotal score indicates the severity of incongruity of the joint, and is a limited indication of subluxation. The remaining components of the overall total score are all measures of osteophytic development which, whilst a positive indication of CHD status, is age-dependent (Smith et al. 2006). For the success of any screening using a single test the identification of true positives and the elimination of false negatives is paramount (accuracy). It has been observed that dogs scored at 1 year of age are less likely to have scores for osteophytic development than older dogs (AJ Worth, unpublished obs.), which means that at this age the subtotal score is the only indication of CHD status. If the subtotal score does not detect all dogs with laxity then the predictive value of a hip-extended radiograph for DJD will be lower than a distraction-radiographic method. However, if the subtotal score relates closely to passive laxity then recommendations based on the subtotal score can be used to improve the selection of breeding animals. The aim of this study was to determine the agreement between the NZVA scoring system and the PennHIP distraction index in a tightly controlled population. We hypothesised that the subtotal score, or one or more of its components, would correlate with the distraction index as measured by the PennHIP method, and further that in older dogs, the subtotal score would be in better agreement with the distraction index than the total score.

Materials and methods

German Shepherd police dogs presented to the Massey University Veterinary Teaching Hospital between November 2003 and September 2007 for PennHIP radiographic evaluation and older than the 1 year minimum for NZVA scoring were included in the study. In 37/47 dogs, radiographs were taken for both PennHIP and NZVA assessment on the same day. In a further 10 dogs, the NZVA hip score had been previously determined 4–61 (median 37) months prior to radiographs for PennHIP assessment (older existing breeding stock).

All dogs were deeply sedated with a combination of 0.005 mg/kg medetomidine (Domitor; Pfizer Australia Pty Ltd, West Ryde, NSW, Australia) and 0.1 mg/kg butorphanol (Butorphic; Lloyd Laboratories, Shenandoah IA, USA) I/V, or a combination of
0.01 mg/kg medetomidine and 0.1 mg/kg butorphanol I/M. A small number of dogs required additional doses of 0.005 mg/kg medetomidine I/V to obtain a sufficient depth of anaesthesia for manipulation. The standard NZVA/OFA view was taken with the dogs positioned in dorsal recumbency, then each dog was repositioned for the compression and distraction views as per the PennHIP method (Smith et al. 1990). At the end of the procedure the dogs were given 0.025 mg/kg atipamezole HCl (Antisedan; Pfizer Australia Pty Ltd), to reverse the effects of anaesthesia.

In each case, a single hip-extended ventrodorsal view was submitted to the NZVA CHD Panel (Wellington, NZ), and the PennHIP views were submitted to the PennHIP Analysis Center (Malvern PA, USA). Panellists scoring the radiographs were unaware of the aims of the study, and read the films as normal submissions.

The NZVA CHD scheme measures Norberg’s angle, with \( \geq 105^\circ \) receiving a score of 0, as shown in Figure 1. Deviation from 105° is graded from 1 to 6 (see Table 1). The subluxation score, also referred to as femoral head coverage, is scored by assessing the position of the centre of the femoral head in relation to the dorsal acetabular edge (Figure 1), and is graded 0 to 6 (Table 1). The cranial acetabular edge is graded 0 to 6, according to the descriptions presented in Table 1. The PennHIP method reports a distraction index as a unit-less ratio calculated from the degree to which the centre of the femoral head is lateralisised from the centre of the acetabulum when forcibly distracted by way of a fulcrum, divided by the radius of the femoral head (Figure 2).

**Statistical analysis**

Results from the two schemes were tabulated for statistical analysis. The frequency distribution of individual hip-distraction index scores and NZVA total and subtotal scores were tested for normality using the Kolmogorov-Smirnov test. The relationships between these scores, and with the Norberg’s angle and subluxation-score components of the NZVA score (which were both ordinal measures), were assessed using correlation. Pearson’s correlation

![Figure 1. Ventrodorsal hip-extended radiograph of a German Shepherd dog, showing measurement of subluxation score (right hip) and Norberg’s angle (left hip). Subluxation score (also referred to as femoral head coverage) is determined by the area of the femoral head imaged within the acetabulum (lined region), as determined by the position of the centre of the femoral head relative to the dorsal acetabular rim (dotted line). Norberg’s angle is the angle included between a line drawn connecting the centres of the femoral heads and the effective cranial margin of the acetabulum.](image1)

![Figure 2. Distraction radiograph of a German Shepherd dog, showing the University of Pennsylvania Hip Improvement Program Distraction Index (DI), measured as the distance between the centres of the acetabulum and femoral head (x), divided by the radius of the femoral head (r), when a distraction force is applied to an anaesthetised dog, with the hindlimbs perpendicular to the spine, i.e. \( DI = x/r \).](image2)

<table>
<thead>
<tr>
<th>Score</th>
<th>Norberg’s angle</th>
<th>Subluxation</th>
<th>Cranial acetabular edge (CAE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>( \geq 105^\circ )</td>
<td>‘None’: FHC well centred within the acetabulum</td>
<td>‘Normal’: Even, smooth curve of CAE parallel to the femoral head throughout</td>
</tr>
<tr>
<td>1</td>
<td>100–104°</td>
<td>‘Slight’: FHC lies within DAE, and the medial joint space increases slightly</td>
<td>CAE, as traced laterally, shows flattening, with only the lateral ¼ affected</td>
</tr>
<tr>
<td>2</td>
<td>95–99°</td>
<td>FHC overlies the DAE, and the medial joint space increases obviously</td>
<td>CAE flattened throughout most of its length</td>
</tr>
<tr>
<td>3</td>
<td>90–94°</td>
<td>‘Moderate’: FHC lies just lateral to the DAE, and approximately ½ the femoral head overlies the DAE</td>
<td>CAE is very flat, with slight cranial slope laterally (= slight bi-labiation)</td>
</tr>
<tr>
<td>4</td>
<td>85–89°</td>
<td>FHC is clearly lateral to the DAE, and ¼ of the femoral head overlies the DAE</td>
<td>CAE flat, and moderate bi-labiation</td>
</tr>
<tr>
<td>5</td>
<td>80–84°</td>
<td>‘Gross’: FHC is well lateral to the DAE but the femoral head touches the DAE</td>
<td>‘S’ deformity of CAE of moderate degree, with gross bi-labiation</td>
</tr>
<tr>
<td>6</td>
<td>( \leq 79^\circ )</td>
<td>Total pathological luxation</td>
<td>Entire CAE sloped cranially</td>
</tr>
</tbody>
</table>

FHC = femoral head centre; DAE = dorsal acetabular edge
was used when both parameters were normally distributed; Kendall's $\tau$–b was used when one or both parameters were not normally distributed. This analysis was then repeated using mean distraction index and mean NZVA scores for each dog, to compensate for obliquity that affects the hip-extended view. For all correlations the 95% confidence limits were estimated using the asymptotic standard errors, calculated assuming the null hypothesis. Correlations were considered significant if the 95% CI did not include 0.

The relationship between hip-distraction index and the Norberg's angle and subluxation-score components of the NZVA subtotal score was then assessed using univariate ANOVA, with distraction index as the dependent variable, and right or left hip and either Norberg's angle score or subluxation score as fixed effects. For this ANOVA, to ensure at least six members in each group, Norberg's angle scores were divided into two groups (original score 0 or $\geq 1$), as were scores for subluxation (original score $\leq 2$ or $\geq 3$). These analyses were then repeated, using mean distraction index, Norberg's angle score and subluxation score. For this ANOVA Norberg's angle scores were divided into three groups (original score 0, 0.5, 1), while subluxation scores were divided into four groups (original score $\leq 1$, 1.5, 2, 2.5).

The scores were then dichotomised into either low risk of CHD or increased risk. The threshold for distraction index was set at $>0.3$, previously established as the threshold for susceptibility for DJD (Smith et al. 1993). For the NZVA score three thresholds were used. The first was a total sum of $>2$ in all categories (total score), the second was a subtotal score of $>2$, and the third was a score of $>1$ in any individual category. The sign test was then used to determine whether the three NZVA thresholds identified the same proportion of ‘at-risk’ dogs as the distraction-index threshold. For those thresholds where no significant difference in proportion of ‘at-risk’ dogs was identified the Kappa test was used to estimate the agreement between them. The 95% CI for Kappa were estimated using the asymptotic standard errors, calculated assuming the null hypothesis.

This process was repeated to compare a distraction-index threshold of $>0.5$ with a total score of 8 in the subtotal and total NZVA scores, less restrictive levels that have been customarily used for breeding recommendations. All analyses were undertaken using SPSS 16.0 for Windows (SPSS Inc, Chicago IL, USA).

The distraction indices for the left and right hips in the same dog were significantly correlated (Pearson’s $r=0.5$; 95% CI=0.28–0.73). However, for the NZVA scores there was no significant correlation, with Pearson’s $r=0.12$ (95% CI=–0.08 to 0.32) for the subtotal score, and Kendall’s $\tau=0.07$ (95% CI=–0.15 to 0.28) for total score. This was also true for subluxation score (Kendall’s $\tau=–0.09$; 95% CI=–0.34 to 0.16) but for Norberg’s angle the relationship between the scores of the two hips was significant at the 5% level (Kendall’s $\tau=0.34$; 95% CI=0.07–0.6).

Comparing individual hip scores, the subtotal score and total score for each hip using the NZVA scoring system were significantly correlated (Pearson’s $r=0.94$ for right hip; 95% CI=0.89–1; Kendall’s $\tau=0.87$ for left hip; 95% CI=0.76–0.99). Additionally, the left distraction index was significantly correlated with both left-subtotal and left-total NZVA scores (Pearson’s $r=0.53$ for subtotal score; 95% CI=0.31–0.75; Kendall’s $\tau=0.29$ for total score; 95% CI=0.02–0.55, respectively), however for right-hip scores there was no such correlation (Pearson’s $r=0.24$ for subtotal score; 95% CI=–0.09 to 0.50; Pearson’s $r=0.24$ for total score; 95% CI=–0.07 to 0.58). This is illustrated in Figure 3. Within each hip, the Norberg’s angle and subluxation score were significantly correlated (Kendall’s $\tau=0.47$ for left hip; 95% CI=0.22–0.72; Kendall’s $\tau=0.37$ for right hip; 95% CI=0.22–0.72). However, when these parameters were compared with the distraction index of the same hip the only significant association at the 5% level was that between the subluxation score and distraction index of the left-hip (Kendall’s $\tau=0.35$; 95% CI=0.16–0.54). For the left-hip Norberg’s angle the association was almost significant (Kendall’s $\tau=0.24$; 95% CI=–0.07 to 0.49) but for the right-hip neither subluxation score nor Norberg’s angle were significantly correlated (Kendall’s $\tau=0.08$; 95% CI=–0.18 to 0.33; Kendall’s $\tau=0.09$; 95% CI=–0.16 to 0.34, respectively).

For individual hips the mean distraction index of hips with a Norberg’s angle score of 0 tended to be lower than those with a score of $\geq 1$ (p=0.063). For the subluxation score, hips that had a score of $\leq 1$ had a lower mean distraction index than those with a score of $\geq 2$ (p=0.025). The relationship between Norberg’s angle score, subluxation score and distraction index for individual hips is illustrated in Figure 4. The effect of right or left hip, or their interaction, was not significant for any of the ANOVA (p=0.1). When mean rather than individual-hip Norberg’s angle score was used there was no association with mean distraction index (p=0.18).

Results

The 47 dogs, 16 male and 31 female, ranged in age from 12 to 84 (median 17.5) months, and weighed 25.5–43.2 (median 30.5) kg. All dogs were being evaluated for the NZ Police dog-breeding programme. Most dogs were New Zealand-bred but a minority of imported lines were represented.

In this group of dogs, the recorded distraction index of individual hips ranged from 0.21 to 0.65 (median 0.36). The median distraction index of the worst (most lax) hip for each case was 0.39. The NZVA subtotal score ranged from 2 to 11 (median 6), and the NZVA total score ranged from 2 to 12 (median 6). Individual hip and mean distraction index and NZVA subtotal scores were normally distributed, as were mean NZVA total scores, and the NZVA total score for the right hip. However, the NZVA total score for the left hip was not normally distributed (p=0.015).

The relationship between the scores of the two hips was significant at the 5% level (Kendall’s $\tau=0.34$; 95% CI=0.07–0.6).
but there was a tendency for it to be associated with the mean subluxation score (p=0.06).

The proportion of dogs identified as being ‘at risk’ of CHD by distraction index, subtotal score, total score and individual-category score were similar for all comparisons when combined results from both hips were used (p>0.5 for all comparisons). However, when results from individual hips were compared there were marked differences between left and right hips. For the left hip all comparisons identified a similar proportion of high-risk hips (p=0.45), but for the right hip all comparisons were different (p<0.04). Reducing the ‘at-risk’ threshold to >1 for the subtotal and total scores resulted in comparisons of the right hip being similar (p>0.17). Using this lower threshold for data for the left hip reduced the p-values for the comparisons but they remained above the 10% level. This threshold was thus used for further comparisons using data from individual hips. The distribution of ‘at-risk’ dogs for each threshold is summarised in Table 2.

The Kappa results for estimation of the agreement between the dichotomised values are shown in Table 3. When a distraction-index score of >0.3 in either hip was compared with total NZVA scores from both hips no agreement was found (Kappa value <0). The Kappa values were higher when results for individual hips were scored, however the values ranged from 0.12 to 0.44, which indicates only a low to moderate level of association.

Using a threshold of >2 for the NZVA total score and >0.3 for the PennHIP index for classification of high-risk dogs, 44/47 (91%) were scored high risk by both methods; however, neither of the two dogs classified as low risk by the NZVA score was scored low risk by the distraction index, and none of the three dogs classified as low risk by the distraction index was scored low risk by the NZVA score (Table 4).

Increasing the distraction-index threshold to 0.5 for either hip identified an appreciably smaller proportion (5/47) of dogs as being at high risk of CHD. However, the proportion of dogs identified using an NZVA threshold of ≥8 was higher for both subtotal (11/47) and total (13/47) score (p=0.07 and 0.021, respectively). Increasing the NZVA threshold to ≥9 resulted in a smaller proportion of dogs identified as at high risk (5/47 and 10/47 for subtotal and total scores, respectively), which for both scores was not significantly different from that identified using the distraction-index threshold of >0.5. The Kappa values using the ≥9 NZVA and 0.5 distraction-index thresholds were moderate for subtotal (0.55; 95% CI=0.16–0.94) and low for total (0.31; 95% CI=0.02 to 0.63) scores.

Table 2. Frequency distribution of 47 German Shepherd dogs classified as high or low risk for canine hip dysplasia (CHD) as determined by one of four thresholds, using the University of Pennsylvania Hip Improvement Program Distraction Index or the New Zealand Veterinary Association (NZVA) Hip Dysplasia Scoring System.

<table>
<thead>
<tr>
<th>Threshold</th>
<th>High risk</th>
<th>Low risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distraction index&lt;sup&gt;a&lt;/sup&gt;</td>
<td>44</td>
<td>3</td>
</tr>
<tr>
<td>NZVA subtotal score&lt;sup&gt;b&lt;/sup&gt;</td>
<td>45</td>
<td>2</td>
</tr>
<tr>
<td>NZVA total score&lt;sup&gt;b&lt;/sup&gt;</td>
<td>45</td>
<td>2</td>
</tr>
<tr>
<td>Categorised NZVA score&lt;sup&gt;c&lt;/sup&gt;</td>
<td>41</td>
<td>6</td>
</tr>
</tbody>
</table>

<sup>a</sup> Score >0.3 in either hip indicative of high risk of CHD
<sup>b</sup> Score >2 indicative of high risk for combined score of both hips
<sup>c</sup> Score >1 in any category of Norberg’s angle, subluxation, cranial acetabular edge, or either hip indicative of high risk of CHD

Table 3. Comparison of Kappa values for agreement between a University of Pennsylvania Hip Improvement Program Distraction Index and the New Zealand Veterinary Association (NZVA) Hip Dysplasia Scoring System, based on dichotomising the score into high and low risk of canine hip dysplasia (CHD). The main figure is a Kappa value, with estimated 95% CI in brackets. Kappa values in bold were significantly greater than 0 (p<0.05).

<table>
<thead>
<tr>
<th>Distraction index</th>
<th>Total NZVA score&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Total NZVA subluxation score&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Categorised NZVA score&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Worst&lt;sup&gt;c&lt;/sup&gt;</td>
<td>−0.05 (−0.10 to 0.00)</td>
<td>−0.05 (−0.10 to 0.00)</td>
<td>−0.10 (−0.18 to 0.02)</td>
</tr>
<tr>
<td>Left hip&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.21 (−0.09 to 0.50)</td>
<td>0.27 (−0.05 to 0.59)</td>
<td>0.22 (−0.08 to 0.53)</td>
</tr>
<tr>
<td>Right hip&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.39 (0.06−0.74)</td>
<td>0.44 (0.10−0.78)</td>
<td>0.12 (−0.11 to 0.34)</td>
</tr>
</tbody>
</table>

<sup>a</sup> Score >2 for total score or >1 for individual hip score indicative of high risk of CHD
<sup>b</sup> Score >1 in any category indicative of high risk of CHD
<sup>c</sup> A distraction-index score of >0.3 in either hip indicative of high risk of CHD
<sup>d</sup> Score >0.3 indicative of high risk of CHD

Table 4. Comparison of the classification of 47 German Shepherd dogs into high and low risk of canine hip dysplasia (CHD), based on the University of Pennsylvania Hip Improvement Program Distraction Index and the New Zealand Veterinary Association (NZVA) Hip Dysplasia Scoring System total score.

<table>
<thead>
<tr>
<th>Distraction index&lt;sup&gt;b&lt;/sup&gt;</th>
<th>High risk</th>
<th>Low risk</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>High risk</td>
<td>42</td>
<td>2</td>
<td>44</td>
</tr>
<tr>
<td>Low risk</td>
<td>3</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>45</td>
<td>2</td>
<td>47</td>
</tr>
</tbody>
</table>

<sup>a</sup> Score >2 indicative of high risk of CHD
<sup>b</sup> Score >0.3 in either hip indicative of high risk of CHD
Laxity of the hip joint has been shown to be a primary phenotypic factor in predicting a dog's susceptibility for developing DJD (Smith et al. 1993, 1995; Popovitch et al. 1995). The PennHIP method was designed to objectively measure passive hip laxity as well as identify osteoarthritis. A recent report using a linear animal model showed the PennHIP index to have a heritability of 0.83 ± 0.11 in a sample of 215 Estrela Mountain dogs (Ginja et al. 2008b). The division of the NZVA score into the subtotal and overall score was intended to provide breeders with better guidance as to the primary phenotypic markers of CHD. Those markers pertaining to laxity and incongruitu, i.e. Norberg's angle, subluxation, and cranial acetabular edge, were grouped separately from those that related to osteophytic development and the presence of DJD.

Our analysis indicated that in the group of animals studied, there was only limited agreement between either the NZVA subtotal or total score and the PennHIP distraction index. The distraction index has been shown to be related to the risk of developing DJD (higher score = higher risk) (Popovitch et al. 1995; Smith et al. 1995). The lack of correlation with the distraction index suggests that increased NZVA subtotal scores above the chosen threshold may not be similarly linked to an increased risk of clinical CHD. Of the parameters subjectively evaluated by the OFA from a hip-extended radiograph none has been shown to be statistically significant as a predictor of DJD (Smith et al. 1993, 1995; Popovitch et al. 1995).

There was better agreement between the distraction indices than the NZVA subtotal scores for the left and right hips of individual dogs. CHD is known to be a bilateral disease (Smith et al. 1990), and the results for the PennHIP indices are more consistent with biological plausibility, i.e. a dog with one bad hip is more likely to have a second bad hip. The disparity between the NZVA subtotal scores for the left and right hip of each dog may be due to the effect of positioning. In order for the dorsal acetabular rim to project equally over each femoral head, the patient must be positioned symmetrically beneath the X-ray beam. Any tilting of the pelvis results in underestimation of the Norberg's angle and subluxation score of one hip, and overestimation of the other. For this reason we also compared the means of the Norberg's angle and subluxation score with the mean distraction index. Better agreement was found between the distraction index and subluxation score than between the distraction index and the score for Norberg's angle (using means). Whilst the mean subluxation score may be a more representative criterion of laxity, in this dataset the improved agreement may have been due to good correlation of the left hip masking the effect of poor correlation of the right hip. Veterinarians should carefully assess symmetry of the obturator foramina and ilial wings prior to submitting hip-extended radiographs to the NZVA Hip Dysplasia Scoring System. Radiography should be repeated until optimal symmetry is obtained, to avoid misinterpretation of subluxation. The radiographs analysed in this study were of high quality and accepted as suitable for scoring by the NZVA CHD panel; despite this it is improbable that they were all exactly symmetrical projections.

The NZVA subtotal and total scores were strongly correlated. The development of DJD in CHD-susceptible dogs is linear into later life, rather than bimodal as previously thought (Smith et al. 2006). The total score includes those criteria that score the presence of DJD. Thus, older dogs (with CHD) would be expected to have greater disparity between total and subtotal scores than a similarly ‘at-risk’ dog radiographed at a younger age. In this population, 24/47 dogs were ≤18 months of age, and could be expected to have lower scores for the DJD components. Intuitively, when changes indicative of DJD are absent in young dogs, CHD status has to be predominantly determined from the components of the subtotal score. However, there was no association between laxity as assessed using the PennHIP index and Norberg's angle, as assessed using the NZVA scheme, and therefore the latter's use as a selection tool is debatable. A significant correlation between Norberg's angle and distraction index has been shown in German Shepherd dogs (Culp et al. 2006) and Estrela Mountain dogs (Ginja et al. 2008a). Those workers compared the actual Norberg's angle measured rather than a grade. As the NZVA CHD system only reports the grade, not the angle measured, it is possible that if Norberg's angle were compared as a continuous variable rather than an ordinal scale the correlation would improve. The accepted threshold for Norberg's angle for normal hip conformation as used by the NZVA is 105° (Lawson 1963) but this has recently been questioned. In a study of Borzois, 42% of those tested had a Norberg's angle of 99–105° (Culp et al. 2006). According to the NZVA scheme these dogs would have received a Grade 1 per hip for Norberg's angle. The Borzoi is considered a non-CHD-susceptible breed, and in the same study all individuals had a distraction index of ≤0.32. Dogs with a distraction index ≤0.3 are considered to have normal (not lax) hips and are thus DJD-non-susceptible (Smith et al. 1993, 1995; Popovitch et al. 1995; Kealy et al. 1997). Using a cut-off of 105° for Norberg's angle may therefore be increasing the false-positive rate, i.e. lowering specificity.

In our study, there was a significant association between the distraction index and subluxation score but the overlap was large and the difference between means was small (0.02 on the PennHIP scale). A hip-extended index taken from an OFA view has been compared with the distraction index using the standard PennHIP method (Kapatkin et al. 2004). The hip-extended index would be similar to the NZVA subluxation score except that the hip-extended index is a unit-less ratio and the NZVA subluxation score is an ordinal scale. There was a significant correlation between a hip-extended index and distraction index (r=0.52) in six large breeds traditionally susceptible to CHD (American Bulldog, Australian Shepherd dog, German Shepherd dog, Golden Retriever, Labrador Retriever, and Rottweiler), suggesting that both measures key on laxity (Kapatkin et al. 2004). However, those authors showed that if the threshold of normality was set at a hip-extended index ≤0.24 (the highest value found in Borzois in the study) then most dogs of other breeds included in the study would have been classified as non-susceptible for DJD. From these data it can be extrapolated that the subluxation score, whilst specific for laxity and a good indicator of laxity if high, is not as sensitive as the distraction index.

The analysis undertaken for this study compared the two tests without assuming that either was the gold standard. All dogs were free of significant lameness at the time of examination, and longitudinal clinical data were not available. End-of-life and necropsy data would be required to conclusively classify dogs as dysplastic or non-dysplastic. We assumed that if the tests were equally effective as screening tools then the data should have shown a
significant degree of agreement between the two tests; this was not evident. Two thresholds were used for the distraction index and NZVA score to independently categorise the dogs as ‘at risk’. In both cases, the proportion of high-risk dogs was very similar irrespective of the threshold. However, agreement between the two systems was poor, as they identified a different group of dogs as low risk when the lower threshold was used and a different group of high-risk dogs when the higher thresholds were used. This meant that for the lower thresholds there was no agreement between either the subtotal or total NZVA scores and the distraction index. It is likely that the absence of dogs that were low risk for both systems may have skewed this analysis, and a larger dataset would be needed to confirm this complete lack of agreement. However, in this dataset, using results from individual hips only increased the agreement to moderate, suggesting that even if a larger population were used agreement would still be insufficient for the two systems to be used interchangeably. Using a higher threshold improved agreement but there was still a considerable number of dogs that were low risk with one test and high risk with the other.

For the standard thresholds the high proportion of dogs identified as ‘at risk’ meant that we gained an accurate assessment of how likely we were to identify a dog as at low risk of CHD using the distraction index or NZVA score when the other test would suggest that it was high risk, but not how likely we were to call a dog low risk with one test when the other suggested it was not. This can be illustrated using the data shown in Table 4. Forty-four dogs had a distraction index >0.3, of which two (5%) had an NZVA score of ±2. Adding another datapoint, changing this figure to 3/45 (7%), or 2/45 (4%), would not markedly change this percentage. In comparison, three dogs had a distraction index ≤0.3, all of which had an NZVA score >2. Adding one datapoint to this figure to either 0/4 or 1/4 would appreciably affect the relationship. In the study by Culp et al. (2006), which analysed 350 clinically normal dogs of seven breeds, the Norberg’s angle had a positive predictive value (proportion of animals testing positive that are truly DJD-susceptible) of only 64% in German Shepherd dogs. It is important to recognise that when such extreme ordinal recommendations are created (subtotal score of ±2 = ‘acceptable’ in the NZVA scheme) it is difficult, if not impossible, to apply meaningful selection pressure. That is, to breed dogs with hips better than they are presently dogs would have to be selected only from the pool of dogs having a score of 1, eliminating 46/48 of the sample as potential breeding stock.

Our interpretation of the results must be viewed in light of the following caveats. The dataset represents a small population of dogs, all of one breed. The New Zealand Police dog-breeding programme has used CHD screening for several years, and the requirement for a high level of activity creates a high selection bias towards sound hip conformation. Despite the low level of clinical disease seen in these lines the phenotypic expression of CHD is widespread according to either screening tool. The general population of German Shepherd dogs may differ from the sample used in the study presented here, due to selection pressures for coat, conformation and temperament. Comparison with the entire population of all dogs of all breeds can be inferred but not guaranteed.

Prior to the development of lameness and demonstrable cartilage injury of the coxofemoral joint no single method of evaluation has proven to be 100% accurate in determining the hip status of dogs (Todhunter and Lust 2003). Cartilage damage precedes osteophytic development, and the clinical signs often correlate poorly with radiographic signs of DJD. It has recently been shown that the development of osteophytes, the hallmark of CHD, is progressive, and that traditional hip scores are dependent on the age at assessment, increasing with time for individual dogs (Smith et al. 2006). The OFA system mandates that dogs are >2 years of age when presented for assessment, based on reports of poor reliability when performed in younger dogs (Corley et al. 1997; Ohlert et al. 2003). The reliability of OFA hip-grade phenotype performed in dogs <2 years of age, i.e. preliminary, has been compared with later evaluations performed after 2 years of age (Corley et al. 1997). Reliability of preliminary evaluations increased as age at the time of preliminary evaluation increased, regardless of whether dogs received a preliminary evaluation of normal hip conformation or CHD. For normal hip conformation the reliability was 89.6% at 3–6 months, 93.8% at 7–12 months, and 95.2% at 13–18 months. Dogs scored in New Zealand are given final grades at any age beyond 12 months, therefore it is likely that up to 5% of dogs would have shifted grade by 24 months. Increasing the age at scoring in New Zealand to 2 years of age, such as occurs in the OFA system, would increase the accuracy of the hip-extended view as a test. Further improvement would be gained by assessment at 5 years. Understandably, the desire from breeders to identify suitable breeding stock prior to 24 months has dictated the age of scoring for the current scheme. The PennHIP index offers a major advantage in that it has shown to be reliable from 4 months of age, i.e. its accuracy as a test is not affected by age from that point on (Smith et al. 1993, 1995; Adams et al. 1998). As a test it is more sensitive but less specific than the NZVA or similar schemes based on the presence of radiographic evidence of subluxation and development of osteophytes. Breed-specific information as to the correlation between distraction index and probability of DJD can be used to aid breeding decisions. The arthritic response to laxity has been shown to differ according to breed; German Shepherd dogs were 6.5 times more likely to develop DJD than Rottweilers with the same distraction-index scores (Smith et al. 2001). The PennHIP Analysis Center calculates an individual dog’s worst distraction index as a point along the breed’s distribution (all those scored). For instance, a German Shepherd dog with a distraction index of 0.33 represents the 75% percentile, 0.40 is at the 50th percentile, and a distraction index of 0.52 is at the 25% percentile, i.e. 75% of dogs have better (= tighter) hips.

Breeders can expect to make greater gains over fewer generations by selecting stock from animals with a low distraction index. The lack of correlation between the NZVA total and subtotal scores and the PennHIP distraction index is of concern as ranking dogs by each method gave disparate results. Given the high reported heritability of the distraction index, the PennHIP index should be considered the best current standard against which other methods should be compared. Selecting for a low NZVA score in young animals before opportunity for the advent of DJD may not successfully reduce the CHD phenotype. The advantage of a hip-extended ventrodorsal view is its low cost and widespread availability, but comparisons between individuals may not be accurate due to poor sensitivity for laxity and false negatives due to positioning artefact. The high heritability of the distraction index found in previous studies suggests that the PennHIP index may be a better selection tool than NZVA/OFA screening when CHD is present within a population (Smith 1997). It has been suggested that due
to its greater sensitivity for laxity the distraction index should be used for selecting breeding lines, whereas the high positive predictive value of the standard hip-extended ventrodorsal view, when DJD is present, is more appropriate for eliminating only those likely to be affected with clinical disease (Ohlerth et al. 2003; Smith et al. 2006).

Declaration of interest

Andrew Worth is the current Convenor of the NZVA Hip and Elbow Dysplasia Scheme, and receives an honorarium for this position. He is also a PennHIP certified veterinarian whose training was supported by the New Zealand Police.

Acknowledgements

The authors would like to acknowledge the radiology service of the Massey University Veterinary Teaching Hospital, and Sergeant Mark Sandford of the New Zealand Police. Funding for this project was granted by Inspector Brendon Gibson of the Police Dog Service on behalf of the New Zealand Government.

References


'Burbidge HM. Changes proposed for inherited disorder control schemes. VetScript XVI (7), 28–9, 2003


Lawson DD. The radiographic diagnosis of hip dysplasia in the dog. Veterinary Record 75, 445–56, 1963


Smith GK, Gregor TP, Rhodes WH, Biery DN. Coxofemoral joint laxity from distraction radiography and its contemporaneous and prospective correlation with laxity, subjective score and evidence of degenerative joint disease from conventional hip-extended radiography. Journal of the Veterinary Medical Association 54, 1021–42, 1993


Submitted 25 February 2009

Accepted for publication 02 September 2009

*Non-peer-reviewed