



INTEROBSERVER VARIATION IN CANINE HIP DYSPLASIA EVALUATION

Ledecky, V., Hluchy, M., Liptak, T., Kuricova, M.

Small animal clinic, Institute of Surgery, Orthopaedics, Radiology and Reproduction
University of Veterinary Medicine and Pharmacy, Komenskeho 73, 041 81 Kosice
The Slovak Republic

kuricovamuvm@gmail.com

ABSTRACT

Canine hip dysplasia (CHD) is a common disease representing an important problem for many dog breeds worldwide. The screening for CHD and breeding programs have been ongoing for many decades but the incidence of disease have failed to be reduced to the expected level. The early diagnosis of CHD is paramount in order to facilitate the early management strategies and to prevent the breeding of the affected individuals. Generally in this area, the emphasis is placed on the radiographic evaluation process, however this is partly a subjective process suggested to be influenced by the experience of the observers. This study was designed to evaluate the interobserver agreement in CHD evaluation based on the Federation Cynologique International system (FCI system). Ten original radiographs were sent to five different groups of observers, from students to certified veterinarians. They were asked to evaluate the ventro-dorsal radiographs according to the FCI system which is the most common system used in Europe to give the final grades (A, B, C, D, E). The grades were converted to numbers

and the data were analysed using a one-way ANOVA test. The results showed that only in 20% of the cases, the interobserver agreement was statistically higher in the group of the most experienced observers when compared to the less experienced group of observers. This means that the level of experience does not always lead to a higher agreement. This could be a problem of widespread objective evaluations of CHD. In addition, there are several different systems of evaluation used in various countries. It is necessary to understand the intention of dog owners, who when buying a dog may be planning its first breeding. Therefore, the “correct” or “incorrect” assessment of the CHD radiographs may not always result in the elimination of affected individuals. We do not know exactly the situation of the development of the hip in some breeds, because scrutineers are receiving only “negative radiographs” for evaluation. Many owners are very educated about hip and elbow dysplasia. It is a time to tell the scientific truth.

Key words: canine hip dysplasia; dog; FCI; interobserver; screening

INTRODUCTION

Canine hip dysplasia (CHD) is a multifocal disease first described in 1935 when Schnelle published the first radiographic description of this disease as a bilateral congenital subluxation of the coxofemoral joint [13]. Although there are several evaluation systems of CHD in continental Europe, the recommendations of the Federation Cynologique International (FCI) using a five-grade scale from A to E are followed to a large extent [3]. Slow progress in decreasing the incidence of CHD remains a fact [11]. This may be due to several factors such as the low sensitivity in detecting hip joint laxity, high interobserver variations, or degenerative joint disease often is not visible at the age the radiographs are made and therefore breeders continue to use dysplastic dogs for breeding purposes [1], [8], [9].

There have been several studies which evaluated the agreement among radiologists in assigning hip dysplasia grades. One study found that the level of agreement between observers, using a subjective method, was very low [14]. Paster et al. [10] stated that intra- and interobserver variations are significant. Saunders et al. [12] also found a significant difference between radiologists evaluating ventrodorsal or dorsoventral hip-extended views.

In the FCI system for screening CHD, dogs are graded as having hips A, B, C, D, or E. Radiographs should be interpreted and scored by a specialized certified veterinarian, approved by the national kennel club and/or the breed club in which the dog is registered. The scoring system combines the subjective standard hip-extended radiographic evaluation with the Norberg-angle measurement. The five different scores represent the severity of the disease. Dogs with A (no signs of CHD) and B (near-normal hip joints) hips are considered non-dysplastic and these dogs are therefore recommended for use in the breeding process. Dogs with grade C are considered mildly affected and can be used in the breeding program in certain instances, whereas grades D and E are considered clearly dysplastic and such dogs are therefore not considered as breeding material.

The objective of this study was to investigate the interobserver agreements within and between groups of differently experienced observers in determining CHD final grading using the FCI grading system.

MATERIALS AND METHODS

For this study we obtained ten original digital radiographs of dogs from the archives of the Department of Surgery, Orthopaedics, Radiology and Reproduction, University of Veterinary Medicine and Pharmacy (UVMP) in Kosice. All radiographs were taken under a standard intramuscular anaesthesia in a position with extended hind legs, properly marked, and numbered from one to ten. They were sent on CDs to 15 different observers. The Norberg angle (NA) was the only information provided. This was intended to minimize the interobserver disagreement caused by differences in the measurements. The normal NA in hips with no dysplasia must be equal to or greater than 105° according to the FCI classification.

Using the FCI method, dogs were evaluated on the basis of the size of the NA, degree of subluxation, shape and depth of the acetabulum, and signs of secondary joint disease. The minimum age for the radiological screening is 12 months for most breeds and 18 months for large and giant breeds. Dogs should be deeply sedated or anaesthetized to achieve complete muscle relaxation and good diagnostic quality radiographs.

Five different groups of observers with decreasing experience were used:

- Group 1.** Experienced veterinarians certificated for evaluating CHD radiographs.
- Group 2.** Experienced private veterinarians without certification.
- Group 3.** Doctors/Lecturers from UVMP.
- Group 4.** PhD. students from the Department of Surgery, UVMP.
- Group 5.** Students from UVMP.

The observers were asked to evaluate all ten radiographs according to the FCI criteria and to give the final grade (A, B, C, D, E). The grades were converted into numbers to be used in statistical analysis (A-1, B-2, C-3, D-4, E-5). An agreement score was derived based on the FCI score for each radiograph and for each CHD grade. We tested whether the agreement scores differed between the experienced and inexperienced groups of observers and between observers in each group. All data were analysed using one-way ANOVA and $P < 0.05^*$ was considered statistically significant.

RESULTS

There were three observers in each of the 5 groups (Table 1). This table shows the amount of radiographs which each observer assigned to each grade of A, B, C, D, or E. The results showed that the more experienced subgroups graded more radiographs with grade A and the inexperienced subgroups graded more radiographs with grade E.

Comparisons were performed for the average scores (Mean \pm SD) of all ten radiographs for the three members in each group (Table 2). There was no statistical difference in the average grading in the groups 1, 2, 3, 4, and 5 ($P > 0.05$).

Table 2 summarizes the evaluations by the individual observers with the number of radiographs evaluated (N), mean values of CHD grades with the standard deviation (SD) and P value expressing that between the evaluations of individual observers and there was no statistically significant difference ($P > 0.05$).

Although the mean value can be the same in groups 1 and 5, the value of the SD demonstrates low interobserver agreement within the groups. For the more experienced observers, this agreement is slightly higher when compared to

the less experienced groups of observers, which confirmed the hypothesis that agreement is growing with experience.

This can be illustrated also by analysing the pooled standard deviations as a weighted average of each group's standard deviations. Data expressed in Table 3 show that the SD increases with decreased level of experience.

The evaluation of the level of agreement for the 5 groups based on 10 different x-rays expressed the following findings. For radiographs 1, 2, 3, 4, 6, and 7 there was confirmed statistical disagreement by means of the difference with $P < 0.05$ (0.011, 0.027, 0.042, 0.01, 0.015, and 0.00, respectively). For radiograph number 10 the P value was 0.001.

DISCUSSION

The primary goal for the ongoing CHD screening program is to exclude genetically susceptible individuals from the breeding pool. Because CHD is a polygenetic heritable trait [7] and current screening systems rely on interpretation of radiographs, their efficacy in reducing CHD is limited [3], [17]. Despite intensive screening for 4 decades, the

Table 1. Relative agreement between observers

Group of observers	Observer	Grade A [%]	Grade B [%]	Grade C [%]	Grade D [%]	Grade E [%]
1	1.1	–	–	40	40	20
	1.2	–	10	70	10	10
	1.3	10	20	50	–	20
2	2.1	50	20	20	10	–
	2.2	20	20	10	30	20
	2.3	–	60	10	30	–
3	3.1	20	20	10	40	10
	3.2	10	30	20	30	10
	3.3	10	30	–	40	20
4	4.1	10	10	20	20	40
	4.2	10	20	10	30	30
	4.3	–	20	20	10	50
5	5.1	–	20	20	30	30
	5.2	10	20	10	30	30
	5.3	10	10	10	40	30

Table 2. Summary of evaluation by individual observers

Observer	N	Mean	SD	P
1.1	10	3.800	0.789	
1.2	10	3.200	0.789	0.176
1.3	10	3.000	1.227	
2.1	10	2.100	1.287	
2.2	10	3.100	1.524	0.229
2.3	10	2.700	0.949	
3.1	10	3.000	1.414	
3.2	10	3.000	1.247	0.852
3.3	10	3.000	1.418	
4.1	10	3.700	1.418	
4.2	10	3.500	1.434	0.812
4.3	10	3.900	1.287	
5.1	10	3.700	1.160	
5.2	10	3.500	1.434	0.926
5.3	10	3.700	1.337	

Table 3. Values of P in each group and pooled SD

Observer group	P	SD
1	0.176	0.9661
2	0.229	1.275
3	0.852	1.362
4	0.812	1.381
5	0.926	1.315

prevalence of CHD is still as high as 40 % in some breeds [2]. Since dogs judged to have moderate or severe CHD cannot be used for breeding purposes, the screening techniques need to be as uniform as possible with a high intra- and interobserver agreements. However, other studies as well as this article indicate that interobserver agreement is low [15]. We can say, that our results are a consequence of the different levels of experience within the groups of observers. Overall, 70 % of all cases confirmed that the difference in assessing the degree of hip dysplasia is linked to the level of experience.

Because of the impact of age at screening and the use of sedation, not all screening approaches are standardized and interobserver agreements seem to be lower also because there is no gold standard for the diagnosis of CHD. The same hip-extended screening system even differs among countries [16]. When a gold standard is available, sensitivity and specificity can be provided. Moreover, if there are many different clubs for the same breed, the clubs may choose different people to evaluate CHD in the same breed which makes it difficult to have a national standard for determining the CHD statistics within one breed [4].

Disagreement between observers inevitably leads to a considerable number of false-positive (loss of genetic variation) and false-negative dogs (genetically affected). This allows false-negative dogs to breed, maintaining hip dysplasia in the population, whereas false-positive dogs, which could decrease the susceptibility for hip dysplasia, are rejected from the pool. This may explain the slow progress of decreasing hip dysplasia over the past few decades [11].

In our opinion, as well as by other authors, the credibility of the FCI screening method for canine hip dysplasia, using the standard hip-extended radiologic view, as currently applied in most European countries, is questionable. The results of Fortrie et al. [5] also demonstrated that the recognition and presence of the radiologic signs of CHD is highly dependent on the expertise and specialty of the observer. One example to increase the positive impact of screening on dog breeding is in Sweden, where ~50 % of all susceptible breeds are screened annually [6]. This makes it possible to identify the symptoms of a degenerative disease where early stages may be diagnosed by screening at a young age.

CONCLUSIONS

Based on the results achieved, it is clear that there is a need for unification and objectivity of the process of CHD evaluations. Despite years of efforts to eliminate this disease, the incidence remains high, which is currently largely related to the problems in evaluation and grading systems; thus, the screening and examination techniques used by different observers need to be as uniform as possible. There is a clear need for the evaluation of the hip for dysplasia to be changed because radiology is not the optimal method for real confirmation of dysplastic or non-dysplastic dogs.

REFERENCES

1. Brass, W., 1989: Hip-dysplasia in dogs. *J. Small Anim. Pract.*, 30, 166—170.
2. Coopman, F., Verhoeven, G., Saunders, J., Duchateau, L., Van Bree, H., 2008: Prevalence of hip dysplasia, elbow dysplasia and humeral head osteochondrosis in dog breeds in Belgium. *Vet. Rec.*, 163, 654—658.
3. Flückiger, M., Lang, J., Binder, H., Busato, A., Boos J., 1995: The control of hip dysplasia in Switzerland. A retrospect of the past 24 years. *Schweiz. Arch. Tierheilkd.*, 137, 243—250.
4. Flückiger, M., 2007: Scoring radiographs for Canine Hip Dysplasia — The big tree organisations in the world. *Europ. J. Comp. Anim. Pract.*, 17, 135—140.
5. Fortrie, R., Verhoeven, G., Broeckx, B., Duchateau, L., Janssens, L., Samoy, Y., et al., 2015: Intra- and interobserver agreement on radiographic phenotype in the diagnosis of Canine Hip Dysplasia. *Vet. Surg.*, 44, 467—473.
6. Swenson, L., Audell, L., Hedhammar, A., 1997: Prevalence and inheritance of and selection for hip dysplasia in seven breeds of dogs in Sweden and benefit: cost analysis of a screening and control program. *JAVMA*, 210, 207—214.
7. Janutta, V., Hamman, H., Distl, O., 2006: Complex segregation analysis of canine hip dysplasia in German shepherd dogs. *J. Hered.*, 97, 13—20.
8. Keller, G., 2003: *The Use of Health Databases and Selective Breeding: A Guide for Dog and Cat Breeders and Owners*. 4th edn., Orthopaedic Foundation of Animals Inc., Columbia, MO, USA, 80 pp.
9. Morgan, J. P., Wind, A., Davidson, A. T., 2000: Hip dysplasia. In *Hereditary Bone and Joint Diseases in the Dogs*. Schlutersche, Hannover, Germany. 131—171.
10. Paster, E. R., Lafond, E., Biery, D. N., Iriye, A., Gregor, T. P., Shofer, F. S., Smith, G. K., 2005: Estimates of prevalence of hip dysplasia in Golden Retrievers and Rottweilers and the influence of bias on published prevalence figures. *J. Am. Vet. Med. Assoc.*, 226, 387—392.
11. Powers, M. Y., Biery, D. N., Lawler, D. F., Evans, R. H., Shofer, F., Mayhew, P., et al., 2004: Use of the caudolateral curvilinear osteophytes as an early marker for future development of osteoarthritis associated with hip dysplasia in dogs. *J. Am. Vet. Med. Assoc.*, 225, 233—237.
12. Saunders, J. H., Godefroid, T., Snaps, F. R., Francois, A., Farnir, F., Balligand, M., 1999: Comparison of ventrodorsal and dorsoventral radiographic projections for hip dysplasia diagnosis. *Vet. Rec.*, 145, 109—110.
13. Schnelle, G. B., 1935: Congenital subluxation of the coxofemoral joint in a dog. *Univ. Pennsylv. Bulletin*, 15.
14. Smith, G. K., Biery, D. N., Rhodes, W. H., 1996: Between- and within-radiologist accuracy of subjective hip scoring of the ventrodorsal hipextended radiograph (Abstract). In *Proc. Int. Symp. Hip Dysplasia and Osteoarthritis in Dogs*. Cornell University, New York, USA, 20.
15. Verhoeven, G., Coopman, F., Duchateau, L., Saunders, J. H., Van Russen, B., Van Bree, H., 2007: Interobserver agreement in the diagnosis of canine hip dysplasia using the standard ventrodorsal hip-extended radiographic method. *JSAP*, 48, 387—393.
16. Verhoeven, G., Fortrie, R., Van Ryssen, B., Coopman, F., 2012: Worldwide Screening for Canine Hip Dysplasia: Where Are We Now? *Vet. Surg.*, 41, 10—19.
17. Willis, M. B., 1997: A review of the progress in canine hip dysplasia in Britain. *J. Am. Vet. Med. Assoc.*, 210, 1480—1482.

Received March 21, 2016