# Inter-observer agreement on the radiographic assessment of canine hip dysplasia

*Herhaalbaarheid van de radiografische beoordeling van heupdysplasie bij de hond*

**[1,3]F. Coopman, [1]G. Verhoeven, [2]D. Paepe, [1]H. van Bree, [3]L. Duchateau, [1]J.H. Saunders**

[1]Department of Medical Imaging, Faculty of Veterinary Medicine, Ghent University,
Salisburylaan 133, 9820 Merelbeke, Belgium

[2]Department of Medicine and Clinical Biology of Companion Animals, Faculty of Veterinary Medicine, Ghent University,
Salisburylaan 133, 9820 Merelbeke, Belgium

[3]Department of Physiology and Biometrics, Ghent University, Salisburylaan 133, 9820 Merelbeke, Belgium

## ABSTRACT

**In order to gain insight into the process of evaluating the hip quality of potential breeding dogs, we estimated inter-observer agreement on the assessability of radiographs, on the morphological observations and on the final scoring of canine hip dysplasia. One hundred radiographs were assessed in terms of their radiographic quality and the morphological traits of the hips by six individual observers and by two groups of two observers each. These six observers and two groups were also asked to evaluate the radiographs in accordance with the FCI (Fédération Cynologique Internationale) instructions. There was no consistent pattern between observers in the evaluation of the technical quality of the radiographs. For some of the observers there was a small amount of agreement in the evaluation of the morphological characteristics. The agreement between the evaluations of the two groups was greater (harmonization between observers). The level of agreement in final scoring (FCI; A, B, C, D, E) was low between some observers, but acceptable between others. These results suggest that the assessment of the quality of ventrodorsal extended radiographs of hips, the assessment of the morphological traits of hips and the final scoring of the hip quality is highly variable between observers, ranging from total disagreement to nearly full agreement. Solutions must be found for maximizing the agreement between the different observers.**

## SAMENVATTING

Om de manier van het beoordelen van de heupkwaliteit beter te begrijpen, werd de herhaalbaarheid van zowel de technische kwaliteitsbeoordeling van de radiografische opname, als van de verschillende morfologische beoordelingen en van de eindbeoordeling van heupdysplasie bij de hond door verschillende beoordelaars in kaart gebracht. Honderd radiografieën werden technisch gekeurd en de heupen werden morfologisch beoordeeld door zes individuele beoordelaars en door twee groepen van telkens 2 beoordelaars. Deze zes beoordelaars en de twee groepen van telkens 2 beoordelaars werden ook gevraagd deze opnamen volgens de FCI (Fédération Cynologique Internationale) instructies te beoordelen. Het was duidelijk dat de technische kwaliteitsbeoordeling sterk verschilde afhankelijk van de beoordelaar. De overeenkomst was klein tussen de beoordelingen van de morfologische kenmerken voor sommige beoordelaars. De overeenkomst tussen de beoordelingen van de twee groepen was groter (harmonisatie tussen beoordelaars). Er was geen grote overeenkomst tussen de eindscore (FCI; A,B,C,D,E) gegeven door sommige beoordelaars, terwijl de overeenstemming tussen de beoordelingen van andere beoordelaars dan weer van een aanvaardbaar niveau was. De resultaten van deze studie tonen aan dat de kwaliteitsbeoordeling van ventrodorsale gestrekte radiografieën van heupen, het beoordelen van morfologische kenmerken en het toekennen van een eindbeoordeling zeer variabel kunnen zijn afhankelijk van de beoordelaar, gaande van totaal verschillend tot bijna gelijk. Het is noodzakelijk om oplossingen te vinden waardoor de overeenkomst tussen de beoordelingen van de verschillende beoordelaars kan worden vergroot.

## INTRODUCTION

Canine hip dysplasia (CHD) is an abnormal development of the hip joint (Fox *et al.*, 1987). Laxity is increased, hips are partially or totally luxated, the femoral head and acetabulum are abnormal and deformed, and osteo-arthritic changes become apparent. It is a multifactorial disorder with both genetic and environmental factors influencing the outcome of the disease (Lust, 1997).

CHD is a highly prevalent disease (Kapatkin *et al.*, 2002) and is currently present in 19.3 % of the total canine population (Rettenmaier *et al.*, 2002). This rate is comparable with the results in Belgium and Europe (Coopman *et al.*, 2004). Many authors agree that the prevalence of CHD is even underestimated (Tomlinson and McLaughlin, 1996; Smith, 1997; Rettenmaier *et al.*, 2002; Coopman *et al.*, 2004). Therefore, it can be concluded that CHD remains an important problem in breeding dogs and that the current eradication programs are failing to decrease the prevalence of this disorder sufficiently.

To prevent affected dogs from being used for breeding (Morgan *et al.*, 2000), an essential feature of any eradication program is to accurately and reproducibly score the hip quality status. There is much discussion concerning the rate of accuracy and reproducibility of different scoring systems, especially for the systems that are based on the standard ventrodorsal hip extended radiographic method (Smith *et al.*, 1992; 1996; Stur *et al.*, 1996; Saunders *et al.*, 1999; Paster *et al.*, 2005; Keller, 2006; Verhoeven *et al.*, 2007).

Different systems are currently being used for scoring the hip status of dogs on hip extended radiographs (Brass, 1993; Flückiger, 1993; Gibbs, 1997; FCI, 2006; Keller, 2006). In all systems, the final scoring is the result of the combination of subjective evaluations and measurements of morphological traits of the hip joint, such as joint space incongruency, level of subluxation, the shape of the weight-bearing dorsal acetabular rim, presence or absence of osteoarthritis, osteophytes and sclerosis, and the shape of the femoral head and acetabulum (Brass, 1993; Flückiger, 1993; Gibbs, 1997; FCI, 2006; Keller, 2006). To assist in providing the final score, Flückiger (1993) and the British Veterinary Association/ Kennel Club (BVA-KC) system (Gibbs, 1997) use additional observations such as the presence of the curvilinear enthesophyte (Morgan line), the shape of the sclerotic band of the subchondral acetabular bone and the positioning of the femoral head center in relation to the dorsal acetabular edge. In addition, measurable traits such as the Norberg angle (NA) and femoral head coverage are used as references (Brass, 1993; Flückiger, 1993; Gibbs, 1997; FCI, 2006; Keller, 2006). Although in theory these parameters have to be measured by the observers, they are often estimated without measurement.

The final scores used by the FCI are (Brass, 1993; FCI, 2006): “A” (hips of very good to excellent hip quality), “B” (hips of sufficient to good hip quality), “C” (mild hip dysplasia), “D” (moderate hip dysplasia) and “E” (severe hip dysplasia). Traditionally, these FCI scores have been used by the Belgian committee and Belgian breeders. The scores used by the Orthopedic Foundation for Animals (OFA) - (USA) are excellent, good, fair, borderline, mild, moderate and severe hip dysplasia (Keller, 2006). In the United Kingdom, the BVA-KC final score is between 0 and 53 for each hip, with an overall hip quality of 0 to 106. An individual hip quality of 11 is considered to be dysplastic (Gibbs, 1997).

Selection procedures are based on the different final scores and not on the distinction between affected or not. Selection procedures are different between countries and breeds, and often not obligatory. Although A and B are considered to be non-dysplastic, it makes a (financial) difference especially for males to be classified as an “A” and not a “B” dog. Assessing hips to be mildly rather than moderately affected makes a difference in breeding restriction and breeding prohibition. Therefore, to breeders, a correct assessment of hips according to these final scores is often more important than the distinction between dysplastic or not.

The aim of this article is to describe the agreement between different individual observers and between two groups of observers on the assessability of radiographic quality, on different morphological traits and on the final FCI score A, B, C, D and E.

## MATERIALS AND METHODS

### Data collection

One hundred radiographs were collected from the archives of the Belgian National Committee for Inherited Skeletal Disorders (NCISD). Thirty of these radiographs had been classified as “A” hips, thirty as “B” hips, thirty as “C” hips, five as “D” hips and 5 as “E” hips in previous meetings of the NCISD in 1999 and 2000. The relative frequency of the different final scores is a reflection of the reality in the common practice of a committee assessing hips. Many more hips of A, B and C quality are sent for official evaluation. Hips showing moderate (D) to extreme (E) signs of CHD are often not sent to the committee. This is due to the fact that in some breeds, D and E hips are not allowed for breeding and therefore a request for an official breeding permit would serve no purpose. Practitioners are capable of recognizing “D” and “E” cases. As an overall result, observers deal more often with excellent to borderline and mildly affected cases than with moderate to extreme dysplastic cases.

To begin with, six members of the NCISD were asked to evaluate the radiographs of the hips of these 100 dogs individually. The observers were free to accept (as assessable) or to reject (as not assessable) each radiograph for its technical qualities. For the radiographs that were accepted for evaluation, the observers were asked to evaluate the femoral head (normal or abnormal), joint congruency (congruent or incongruent), subluxation (present or not), osteophytes (OP) (present or not), Morgan line (present or not), osteoarthritis (OA) (from none, mild, moderate to severe) and the Norberg Angle (NA) ($\geq 105°$ or $< 105°$) for each hip joint. The decision to measure the NA was left to the discretion of the observers.

Finally, the observers were asked to give a final score according to FCI criteria (A, B, C, D, E; Brass, 1993; FCI, 2006). All six observers had long experience in officially reading hips. Two of them were ECVDI (European College of Veterinary Diagnostic Imaging) approved radiologists and one was ECVS (European College of Veterinary Surgeons) approved (orthopedic surgeon).

Next, the same 100 radiographs were presented to two groups of two observers each. Two observers with more than 25 years of experience in film reading, one on whom was the ECVS approved orthopedic surgeon, were put together in one group, while the two ECVDI approved radiologists composed the other group. The two groups had to state separately whether each radiograph was assessable or not. If both members of a group agreed on the assessability, then they were asked to give an evaluation of the morphological traits of the hip joints and to give a final score (A to E). Immediately after scoring, the results of the two groups were compared. The hips with different individual and final reading results were then presented to the two groups working together. Then the four members had to discuss the differences and were forced to find a consensus on the assessability, the morphological traits and the final score. A post-screening discussion with the four observers of the two groups was held and notes were taken on this discussion and summarized in a report.

No data on age, breed, practitioner or the previous score was provided to the observers. All radiographs were evaluated within the same time frame.

**Data processing and statistical analysis**

All available data were collected and registered in an Excel worksheet. SPSS 11.0 for Windows was used for the statistical analyses (Chi-square; Cohen's kappa coefficient).

A descriptive analysis was performed and the relative frequency of the different evaluations for every observer and group of observers was estimated.

Additionally, the inter-observer agreement (Cohen's kappa coefficient) between pairs of observers and between groups was estimated. For the estimation of the Cohen's kappa coefficient, only radiographs that were technically sound according to all individual observers or according to both groups were considered.

Using the Cohen's kappa coefficients of the pairs of observers, mean, maximum en minimum values were estimated.

All final scores given by the different observers and the two groups were evaluated for each dog individually.

## RESULTS

The evaluation of the technical quality of the radiographs did not show a consistent pattern. One observer considered all radiographs to be assessable. One observer rejected 7 radiographs because of insufficient radiographic quality, two observers rejected 8 radiographs, and the last two observers rejected 16 and 28 radiographs, respectively. These differences were significant ($p < 0.001$). The rejection rate of radiographs differed significantly ($p < 0.001$) between the two groups, as well. The first group rejected 51 radiographs and the second group rejected 27 radiographs. Four radiographs that were rejected by the second group were accepted by group 1, which meant that agreement on technical acceptability existed only for 45 radiographs.

The relative frequencies of the scores of all observers and both groups are listed in Tables 1 and 2. The summarized inter-observer agreements between the six individual observers and the inter-observer agreement between the two groups for the 7 morphological traits and for the final score are listed in Table 3.

All radiographs had at least two final scores given by two different observers. For the cases in which all observers accepted the radiograph, the dog was given 6 final scores. Critical individual dog evaluation revealed that no consensus was achieved between the individual observers for 78 dogs. In 65 cases, disagreement occurred in scores A-B, B-C, C-D or D-E. In 13 cases, disagreement occurred in scores over three different classes (A-B-C, B-C-D and B-D-E). In the group screening, 45 radiographs had two final scores. In 13 cases, there was no consensus on the final score. The differences were a one-class difference, except for one case, with an evaluation of A versus C. One group consistently scored one class lower than the other group.

The post-screening discussion clearly indicated that there were two main streams. A more strict group (ECVDI – radiologists) that consistently rejected radiographs of poor radiographic quality, and another more tolerant group (> 25 year of experience group) that did not reject poor quality radiographs ($n = 16$). The latter group did not reject radiographs in cases for which they were convinced that the hip quality was clearly insufficient (D and E) and that rejection would only be worthwhile if there was doubt on the final score. The more stringent group was not convinced that the presumed D and E cases were indeed true D and E hips. The stringent group also gave less A (= excellent) scores because they were convinced this score should be preserved for the real "Optima Forma" cases, as defined in the FCI criteria (Brass, 1993; FCI, 2006). The more tolerant group explained in the post-screening discussion that they accepted minor hip conformation changes and therefore they gave more "A" scores.

## DISCUSSION

According to our results, the evaluation of radiographs by individuals and by groups leads to differences in the determination of radiographic quality. The discrepancy between rejection rates suggests that at the current time the evaluation of radiographic quality is not a standardized procedure based on objective parameters and that no clear distinction can be made between radiographs that provide sufficient diagnostic information and radiographs that fail to do so. The post-screening discussion between the two groups confirmed this lack of agreement. This dis-

**Table 1. The relative frequencies of the different morphological scores for six individual observers (O) and for two groups of two observers each (G). The number between brackets is the number of radiographs for which the technical quality was considered assessable.**

| | | O1 (100) | O2 (84) | O3 (92) | O4 (92) | O5 (93) | O6 (72) | G1 (49) | G2 (73) |
|---|---|---|---|---|---|---|---|---|---|
| **Femoral head** | Normal | 85.0 | 73.8 | 93.8 | 84.6 | 80.0 | 76.9 | 91.3 | 89.1 |
| | Abnormal | 15.0 | 26.2 | 6.2 | 15.4 | 20.0 | 23.1 | 8.7 | 10.9 |
| **Joint space** | Congruent | 3.0 | 18.8 | 12.5 | 15.6 | 18.8 | 14.1 | 15.2 | 28.3 |
| | Incongruent | 97.0 | 81.2 | 87.5 | 84.4 | 81.2 | 85.9 | 84.8 | 71.7 |
| **Osteoarthritis** | None | 80.0 | 64.5 | 46.8 | 69.4 | 79.0 | - | 84.8 | 86.5 |
| | Mild | 12.0 | 8.1 | 33.8 | 16.1 | 8.1 | - | 10.9 | 8.9 |
| | Moderate | 5.0 | 14.5 | 14.5 | 8.1 | 8.1 | - | 0.0 | 0.0 |
| | Severe | 3.0 | 12.9 | 4.7 | 6.5 | 4.8 | - | 4.3 | 4.6 |
| **Morgan line** | None | 58.0 | 71.2 | 62.1 | 65.2 | 63.6 | 68.2 | 67.4 | 68.4 |
| | present | 42.0 | 28.8 | 37.9 | 34.8 | 36.4 | 31.8 | 32.6 | 31.6 |
| **Norberg Angle** | ≥ 105° | 66.0 | 53.8 | 52.3 | 60.0 | 58.5 | 52.3 | 41.3 | 71.7 |
| | < 105° | 34.0 | 46.2 | 47.7 | 40.0 | 41.5 | 47.7 | 58.7 | 28.3 |
| **Osteophytes** | None | 83.0 | 70.8 | 81.5 | 75.4 | 83.1 | 83.1 | 87.0 | 84.8 |
| | Present | 17.0 | 29.2 | 18.5 | 24.6 | 16.9 | 16.9 | 13.0 | 15.2 |
| **Subluxation** | None | 78.0 | 53.8 | 55.4 | 63.1 | 66.2 | 61.5 | 71.7 | 74.3 |
| | Present | 22.0 | 46.2 | 44.6 | 36.9 | 33.8 | 38.5 | 28.3 | 25.7 |

**Table 2. The relative frequencies of the different final scores according to the FCI regulations (A to E), as evaluated by the six individual observers (O) and by the two groups of two observers each (G). The number between brackets is the number of radiographs that were evaluated.**

| % | O1 (100) | O2 (84) | O3 (92) | O4 (92) | O5 (93) | O6 (72) | G1 (49) | G2 (73) |
|---|---|---|---|---|---|---|---|---|
| A | 8.0 | 27.4 | 16.3 | 19.5 | 25.8 | 11.1 | 12.2 | 24.6 |
| B | 50 | 30.9 | 35.8 | 39.1 | 34.4 | 33.3 | 48.9 | 28.7 |
| C | 26 | 11.9 | 26.1 | 14.1 | 15.0 | 27.8 | 20.4 | 17.8 |
| D | 14.0 | 28.5 | 21.7 | 27.1 | 24.7 | 25.0 | 16.3 | 28.7 |
| E | 2.0 | 1.1 | 0.0 | 0.0 | 0.0 | 2.8 | 2.0 | 0.0 |

**Table 3. Maximum, mean and minimum values and standard deviation (SD) of the 6 different observer agreements (Cohen kappa values) for the shape of the femoral head (FH), congruency of the joint space (JS), level of osteoarthritis (OA), presence of osteophytes (OP), degree of subluxation (SL), the Norberg Angle (NA), the presence or absence of a Morgan line (ML) and the final score (FS). N is the number of radiographs that were considered to be assessable and were scored by all six observers. G represents the K-value between the two groups for the same morphological traits and final scoring on 45 radiographs.**

| Trait | N | Mean | Min. | Max. | SD | Group (G) |
|---|---|---|---|---|---|---|
| Femoral head (FH) | 65 | 0.55 | 0.31 | 0.84 | 0.17 | 0.69 |
| Joint space (JS) | 64 | 0.38 | 0.09 | 0.66 | 0.20 | 0.63 |
| Osteoarthritis (OA) | 62 | 0.40 | 0.22 | 0.59 | 0.11 | 0.84 |
| Osteophytes (OP) | 65 | 0.58 | 0.45 | 0.74 | 0.07 | 0.91 |
| Subluxation (SL) | 65 | 0.72 | 0.46 | 0.93 | 0.14 | 0.79 |
| Norberg Angle (NA) | 65 | 0.74 | 0.66 | 0.88 | 0.06 | 0.35 |
| Morgan line (ML) | 66 | 0.71 | 0.58 | 0.84 | 0.08 | 0.70 |
| Final score (FS) | 65 | 0.50 | 0.27 | 0.70 | 0.12 | 0.62 |

agreement creates a difficult situation for the referring veterinarian in terms of transparency in evaluating radiographic quality by the committee.

The relative frequencies of the morphological traits (Table 1) indicate that variation in scoring of these traits is sometimes high between individual observers but less between the two groups, as is confirmed by the ϰ values in Table 3. This might suggest a harmonization effect when evaluations occur in group.

The level of agreement between observers in final scoring is sometimes very low (Table 2). The fact that only 22 of 100 dogs have a similar final score given by 2 to 6 observers confirms the lack of agreement between observers in final scoring.

In view of the ϰ-values, which are shown in Table 3, and on the basis of the low level of agreement for the final scoring of an individual dog, it can be concluded that the instructions for scoring morphological traits and for giving a final score are too subjective. The post-screening discussion confirms the disagreement between groups of observers.

It is obvious that the credibility of an evaluation procedure showing such discrepancies in rejection rate and final scoring of the same radiographs will be called into question both by practitioners and by breeders. Of course, 100 % agreement is only hypothetical and will never be achieved, but such great discrepancies are simply unacceptable.

This study and the study of Saunders *et al.* (1999) might suggest that the large variation in inter-observer agreement reflects only the Belgian situation, especially because in other countries that use the OFA system (Keller, 2006) or the FCI system (Stur *et al.*, 1996), high inter-observer agreement was found (73.5% - 70 %). However, this result has not been confirmed (Smith *et al.*, 1992; 1996; Paster *et al.*, 2005; Verhoeven *et al.*, 2007). The explanation for the discrepancy between Keller (2006) and the other investigators (Smith *et al.*, 1992; 1996; Paster *et al.*, 2005; Verhoeven *et al.*, 2007) might be found in the results of the present study (Table 3). Between some observers, a Κ-value of > 0.8 was found for some morphological traits and a Κ-value of 0.7 for the final score, indicating that there is high agreement. This is in agreement with the study of Stur *et al.* (1996) and Keller (2006), which indicates that these very good results are correct. However, Smith *et al.* (1996) found low inter-observer agreement scores with Κ-values as low as 0.04. In our study, agreement between some observers was also that low, indicating that the results found by Smith *et al.* (1996) also reflect a realistic situation.

Different reasons can be found to explain why in some cases high agreement is found and in some cases almost no agreement is seen. According to the OFA criteria, the minimum age for the screening of dogs is 2 years, whereas the minimum age for FCI screening is 12 to 18 months. A correct assessment is much easier at an older age (Keller, 2006), which thereby increases the inter-observer agreement. This is mainly because at an older age, osteo-artritis and osteophytes are more likely to be present. In our study, a high inter-observer agreement is seen for both OA and OP between the two groups containing highly trained observers. Finally, the level of experience of the observers might create a substantial difference in inter-observer agreement (Verhoeven *et al.*, 2007).

It would also be interesting to know the intra-observer agreement. If observers read consistently (high intra-observer agreement), but different from one another, then the improvement of inter-observer agreement can be achieved by data modification. If not, then the reading should be standardized to a greater extent in order to improve both reader consistency and inter-observer agreement.

To improve agreement between observers, a system that classifies dogs in a group of affected and a group of disease-free animals can be considered. Borderline cases can be classified in a separate group and be allowed to breed temporarily. A new evaluation at an older age could be worthwhile for dealing with these borderline cases. The classification of hips in only two or three groups results in higher inter-observer agreement (Keller, 2006; Verhoeven *et al.*, 2007).

The use of measurements and strict classification instructions based on these measurements can help to improve the level of agreement between observers, not only in the final scoring but also in the technical assessment of the radiographic quality. In Table 3, the best agreement score is seen for the Norberg Angle (NA) which, unlike all the other traits mentioned, can be measured. A measurable trait should of course be measured (as was done by the two groups) and not estimated, because otherwise inter-observer agreement drops consistently.

Focusing on the scoring of osteoarthritis (= high inter-observer agreement) on subsequent radiographs of the same dog at different ages, (thereby identifying the dogs that develop OA (= affected) and the ones that do not (= disease free)), is another option in the effort to increase credibility. Evaluation by groups rather than by individuals is an interesting option, as well.

## CONCLUSION

The results presented in this study are in agreement with the results reported in the literature (Smith *et al.*, 1992; 1996; Stur *et al.*, 1996; Saunders *et al.*, 1999; Paster *et al.*, 2005; Keller, 2006; Verhoeven *et al.*, 2007) It can be concluded that the assessment of radiographic quality, the assessment of morphological traits and the final scoring of hips on a standard ventrodorsal hip extended radiograph are extremely variable, ranging from total disagreement to nearly full agreement between individual observers and groups of observers. The reason for this high variability and inconsistency is most likely to be the fact that the system is not yet sufficiently standardized and strict.

## REFERENCES

Brass W. (1993). Hüftgelenkdysplasie und Ellbogenkrankkung im Visier der Fédération Cynologique Internationale. *Kleintierpraxis 38,* 191-266.

Coopman F., Paepe D., van Bree H., Saunders J. (2004).

The prevalence and evolution of canine hip dysplasia in Belgium (Abstr). In: *Annual Scientific Conference Proceedings, The European Association of Veterinary Diagnostic Imaging and The European College of Veterinary Diagnostic imaging.* September 2004, Ghent Belgium, p. 68.

FCI (2006). *FCI Hip Dysplasia Conference,* March 2006, Copenhagen.

Flückiger M. (1993). The standardized analysis of radiographs for hip dysplasia in dogs. Objectifying a subjective process. *Kleintierpraxis 38,* 693-702.

Fox S.M., Burns J., Burt J. (1987). Symposium on hip dysplasia; The dysplastic hip: A crippling problem in dogs. *Veterinary Medecine 82,* 683-693.

Gibbs C. (1997). The BVA/KC scoring scheme for control of hip dysplasia: interpretation of criteria. *The Veterinary Record 13,* 275-284.

Kapatkin A.S., Fordyce H.H., Mayhew P., Smith G.K. (2002). Canine Hip Dyplasia: The Disease and Its Diagnosis. *Compendium 24,* 526-535.

Keller G. (2006). Hip dysplasia. In: *The Use of Health Databases and Selective Breeding. Information brochure of the OFA,* p. 19-34.

Lust G. (1997). An overview of the pathogenesis of canine hip dysplasia. *Journal of the American Veterinary Association 210,* 1443-1445.

Morgan J.P., Wind A., Davidson A.T. (2000). Hip Dysplasia. In: Morgan J.P., Wind A., Davidson A.T. (editors). *Hereditary Bone and Joint Diseases in the Dog,* Schlütersche Hannover, p. 131-171.

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. *Journal of the American Veterinary Association 226,* 387-392.

Rettenmaier J.L., Keller G.G., Lattimer J.C., Corley E.A., Ellersieck M.R (2002). Prevalence of canine hip dysplasia in a veterinary teaching hospital population. *Veterinary Radiology and Ultrasound 43,* 313-318.

Saunders J.H., Godefroid T., Snaps F.R., François A., Farnir A., Balligand M. (1999). Comparison of ventrodorsal and dorsoventral radiographic projections for hip dysplasia diagnosis. *The Veterinary Record 145,* 109-110.

Smith G.K., Gregor T.P., Biery D.N. *et al.* (1992). Hip dysplasia diagnosis: a comparison of diagnostic methods and diagnosticians. *Proceedings, International Symposium on Hip Dysplasia and Osteoarthritis in Dogs.* Keystone, Colorado, p. 20.

Smith G.K., Biery D.N., Rhodes W.H. (1996). Between- and within-radiologist accuracy of subjective hip scoring of the ventrodorsal hip-extended radiograph (abstr). In: *Proceedings of the International Symposium on Hip Dysplasia and Osteoarthritis in Dogs,* p.20.

Smith G.K. (1997). Advances in diagnosing canine hip dysplasia. *Journal of the American Veterinary Association 210,* 1451-1457.

Stur I., Köppel E., Schröder K. (1996). Populationsgenetische Aspekte der Hüftgelenksdysplasie (HD) – Diagnostik beim Hund – Bewertung unter Berücksichtigung differierender HD-Befunde. *Wiener Tierärztlichen Monatsschrifte 83,* 91-97.

Tomlinson J., McLaughlin R.Jr. (1996). Symposium on CHD: Diagnosis and medical treatment. Canine hip dysplasia: Developmental factors, clinical signs, and initial examination steps. *Veterinary Medicine 91,* 25-33.

Verhoeven G., Coopman F., Duchateau L., Saunders J., Van Ryssen B., van Bree H. (2007). Inter-observer agreement in the diagnosis of canine hip dysplasia using the standard ventrodorsal hip-extended radiographic method. *Journal of Small Animal Practice 41,* 387-393.