

This is the peer reviewed version of the following article:

Pulkkinen, H, Reunanen, V, Hyytiäinen, H, Junnila, J, Laitinen-Vapaavuori, O, Lappalainen, A. The intra- and intertester repeatability of radiographic elbow incongruity grading is high in chondrodystrophic dog breeds. *Vet Radiol Ultrasound*. 2020; 61: 329– 335.

<https://doi.org/10.1111/vru.12853>,

which has been published in final form at <https://doi.org/10.1111/vru.12853>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

1 **The intra- and inter-tester repeatability of radiographic elbow incongruity grading is high in**
2 **chondrodystrophic dog breeds**

3

4 Pulkkinen HSM¹, Reunanen VLJ¹, Hyytiäinen HK¹, Junnila JJT², Laitinen-Vapaavuori OM¹,
5 Lappalainen AK¹

6

7 ¹Department of Equine and Small Animal Medicine, University of Helsinki Faculty of Veterinary
8 Medicine. Agnes Sjöberginkatu 2, P.O. Box 66, 00014 University of Helsinki, Finland

9 ²4Pharma Ltd, Arkadiankatu 7, 00100 Helsinki, Finland

10

11 Correspondence: Hanna Pulkkinen, Agnes Sjöberginkatu 2, P.O. Box 66, 00014 University of
12 Helsinki, Finland. hanna.pulkkinen@helsinki.fi

13

14 Key words: elbow dysplasia, incongruity, chondrodystrophic dog, reliability, validity

15

16 **Conflict of interest disclosure:** The authors declare that there is no conflict of interest.

17

18 **Previous presentation/publication disclosure:** The findings of this study have not been
19 previously presented or published.

20

21 **EQUATOR network disclosure:** The authors followed the STARD 2015 reporting guidelines as
22 applicable.

23

24 Abbreviations: INC, incongruity grade; PA, proportion of agreement; CI, confidence interval.

25 **Abstract**

26 Elbow incongruity is a form of elbow dysplasia that causes osteoarthritis, pain and lameness, and is
27 common in chondrodystrophic dog breeds. The objective of this retrospective secondary analysis
28 study was to evaluate the intra- and interobserver repeatability of a novel radiographic incongruity
29 grading system for assessing elbow incongruity in three chondrodystrophic dog breeds – the
30 dachshund, Skye terrier and Glen of Imaal terrier. We conducted an observer agreement study that
31 included 220 mediolateral antebrachial radiographs from 110 dogs with the elbow in 90° flexion.
32 The radiographs were independently assessed by three observers at three time points, using a four-
33 stepped grading scale. The proportion of agreement and Kappa coefficient were calculated. Both the
34 intra- and interobserver proportions of agreement were substantial when three grades were required
35 to be identical (0.705–0.777 and 0.609, respectively), and almost perfect for two identical grades
36 (0.991–1.000 and 0.991, respectively). Some differences in repeatability between breeds were
37 noted; specifically, the intraobserver repeatability was higher in the dachshund, and the
38 interobserver repeatability was lower in the Glen of Imaal terrier. Our study showed that the
39 radiographic imaging protocol and INC grading system has high repeatability when assessing elbow
40 incongruity in chondrodystrophic dog breeds.

41

42 **Introduction**

43 Chondrodystrophy is a breed-characterizing trait caused by a *fgf4* retrogene, which affects dog
44 breeds such as the dachshund, Glen of Imaal terrier and Skye terrier.¹ The characteristic short and
45 curved front legs – which are also included in the breed standards of some of these breeds – is
46 caused by blunted growth of the long bones¹. Additionally, premature closure of the distal ulnar
47 growth plate is also recognized as a common finding in some chondrodystrophic breeds.^{2,3} This
48 causes asynchronous growth of the radius and ulna, which is recognized as an etiology for elbow
49 incongruity. This form of elbow dysplasia^{3–5} is common in the Skye terrier, but for most

50 chondrodystrophic breeds the prevalence is unknown.^{2,3} In addition to elbow incongruity, the short
51 ulna may cause the radius to bend craniocaudally and mediolaterally during growth, which is
52 clinically described as multiplanar angular deformity of the antebrachium and valgus deformity of
53 the carpal joint.³⁻⁷ Elbow incongruity is associated with osteoarthritis, pain and lameness during
54 growth and adulthood.^{2,3}

55
56 Currently, the screening protocol used by the International Elbow Working Group (IEWG) and
57 Orthopedic Foundation for Animals (OFA) is not optimal for detecting elbow incongruity in
58 chondrodystrophic breeds.^{8,9} It is only able to detect the most severely incongruent elbows and
59 osteoarthritic changes of the elbows.^{8,10,11} Lappalainen et al. (2016) proposed a novel grading
60 system for elbow incongruity in chondrodystrophic dog breeds.² The radiographic protocol is based
61 on images obtained with a 90° flexion of the elbow, with the x-ray beam centered mid-radius such
62 that the whole antebrachium (including the carpal joint) is visible in the image.² However, the
63 repeatability of this novel grading system has not been tested, although such studies should be
64 carried out to assess the intra- and interobserver agreement. The objective of our study was to
65 evaluate the intra- and interobserver repeatability of the grading system for elbow incongruity in
66 chondrodystrophic dog breeds. We hypothesize that the grading system is repeatable both within
67 and between observers.

68

69 **Methods**

70 This was a retrospective secondary analysis evaluating observer agreement¹² using mediolateral
71 antebrachial radiographs of chondrodystrophic dogs, originally taken for another research project
72 that had been approved by The National Animal Experiment Board in Finland
73 (ESAVI/9184/04.10.07/2014). Dachshunds, Skye terriers and Glen of Imaal terriers of 1–10 years
74 of age and no history of orthopaedic surgery or a condition that would increase the risk of sedation

75 were eligible for the study and were approved by a qualified veterinarian. These images had been
76 acquired for the other research project using the imaging protocol introduced by Lappalainen et al.
77 (2016).² The images were acquired using computed radiography with an automatic exposure
78 detector, imaging plates and reader (FUJIFILM FCR XG-1 CR-IR 346RU, Fuji Photo Film CO.,
79 LTD. 26-30, Nishiazabu 2-chome Minato-ku Tokyo 106-8620, Japan). S-values of 100-300 were
80 targeted to ensure image quality. One of the observers was involved in supervising the acquisition
81 of the radiographs in the original study and was aware of the dogs signalment and history regarding
82 lameness before the anonymization process was performed for the images. The two other observers
83 were only provided anonymized radiographs. The DICOM-images were anonymized and
84 randomized using computerization. The radiographs were independently graded by three observers
85 (A, B and C) who had 12, 1 and 7 years of experience in veterinary radiology, respectively. Three
86 randomized sets of the radiographs were produced for each observer; each observer graded each
87 image three times. The radiographs were assessed with image analysis freeware (OsiriX MD 9.0 or
88 Horos DICOM viewer v. 2.1.1). Magnification and windowing were allowed as necessary to ensure
89 precise measurements. The grading was performed using the 4-stepped grading system (INC0–
90 INC3) described by Lappalainen et al. (2016; Table 1). Prior to the actual grading, the observers
91 familiarized themselves with the method through discussion with each other. Ten randomly chosen
92 radiographs from the set were then given to each evaluator in order to get accustomed to the
93 grading. Finally, the results were discussed to ensure that all observers were confident with the
94 grading protocol. The first actual grading was done two weeks after the familiarization process. To
95 avoid recall bias, the three gradings were carried out over a 12-week time period, with at least two
96 weeks between gradings. Once a grade was decided, second-look revisions for the image were not
97 allowed.

98

99 All the statistical tests were selected and completed by a professional biostatistician (MSc in
100 biostatistics) with over 10 years' experience conducting clinical and non-clinical trials.
101 Intraobserver repeatability was calculated as the proportion of images in which the observer
102 assigned the same INC grade (proportion of agreement) with 95% confidence intervals (CI). The
103 mode of the three ratings for each individual observer was used to calculate the interobserver
104 agreement (with 95% CIs). The interobserver proportion of agreement was calculated for full
105 agreement (all observers agree), and also for pairwise agreements between each pair of observers.
106 Furthermore, the proportion of radiographs where at least two out of three gradings were identical
107 was calculated for both intra- and interobserver agreement. In addition, Fleiss' Kappa coefficients
108 for the three-way agreement within and between the observers were calculated, and weighted
109 Kappa coefficients (with 95 % CIs) were calculated for the pairwise agreement between the
110 observers. When the Kappa values were evaluated, the following ranges were used: less than zero
111 was considered to have less than a chance agreement; 0.01–0.20 was slight agreement; 0.21–0.40
112 was fair agreement; 0.41–0.60 was moderate agreement; 0.61–0.80 was substantial agreement; and
113 0.81–1.00 was almost perfect agreement.¹³ For the purpose of this study, these categories were also
114 applied to the proportion of agreement values. Statistical analyses were performed using
115 commercially available software (SAS[®] System for Windows, version 9.3, SAS Institute Inc., Cary,
116 NC, USA; and R for Windows, version 3.4.2, R Foundation for Statistical Computing, Vienna,
117 Austria).

118

119 **Results**

120 Digital radiographs from 30 dachshunds, 53 Glen of Imaal terriers and 27 Skye terriers were
121 available for evaluation. The images included both the left and right elbow radiographs of each dog,
122 yielding a total of 220 radiographs from the 110 dogs. All INC grades were represented in the
123 radiographs. INC1 was the most frequent grade for each observer at each time point, and INC3 was

124 the most uncommon. Table 2 reports the frequencies of each grade assigned by each observer at
125 each time point.

126

127 The intraobserver proportion of agreement was substantial (0.705–0.777) when the requirement was
128 for three grades to be identical. When the requirement was for two grades to be identical, the
129 agreement was almost perfect (0.991–1.000). The intraobserver Fleiss' Kappa values also suggested
130 substantial agreement (0.623–0.690) for each of the observers when three identical grades were
131 required (Table 3). The interobserver proportion of agreement was substantial (0.609; 95% CI 0.54–
132 0.67) when all three observers were included; if at least two observers were required to agree on the
133 same grade, the proportion was almost perfect (0.991). The interobserver Fleiss' Kappa value was
134 moderate (0.502; 95% CI 0.45–0.56) when all three observers were required to agree, and moderate
135 to substantial (0.568–0.617) for the pairwise analyses between observers (Table 4). An example of
136 an elbow with high intra and interobserver agreement is shown in Figure 1 and an example of an
137 elbow with low agreement is shown in Figure 2.

138

139 When analyzed by breed, the intraobserver proportion of agreement was highest for the dachshund
140 (0.767–0.850) for all observers. The agreement was lower, and roughly similar for the Glen of
141 Imaal terrier (0.651–0.745) and Skye terrier (0.667–0.759; Table 5). The interobserver proportion of
142 agreement by breed was moderate to substantial, ranging from 0.557–0.717. Based on the Fleiss'
143 Kappa coefficient, agreement was moderate for both the Skye terrier (0.528) and dachshund
144 (0.570), but only fair for the Glen of Imaal terrier (0.400; Table 6).

145

146 **Discussion**

147 Our study showed that the novel grading method for assessing elbow incongruity among
148 chondrodystrophic dogs had excellent intraobserver proportion of agreement. When two grades (out

149 of the three grading time points) were required to be identical, two of the observers had a perfect
150 intraobserver agreement, and the third observer reached an almost perfect agreement (Table 3). The
151 intraobserver agreement was substantial even when grades from all three time points were assessed.
152 The interobserver agreement was almost perfect between two observers, and substantial among all
153 three observers.

154

155 Based on our review of the literature, there are no repeatability studies available for comparisons of
156 radiographic grading of elbow dysplasia in dogs. However, according to repeatability studies of
157 another prevalent radiographic screening method – the Fédération Cynologique Internationale (FCI)
158 hip dysplasia scoring system – proportions of agreement have ranged from 46.3 to 71.3%, and
159 Kappa values from 0.46 to 0.76.¹⁴⁻¹⁶ Thus, the interobserver consistency of INC found in our study
160 appears to be higher than what has been reported for hip dysplasia. The high agreement for the INC
161 grading system noted in our study may reflect the objectivity of the measurement-based protocol,
162 compared to the subjective nature of the hip dysplasia measurement and FCI score.

163

164 Regarding the observer-related factors that may contribute to the grading consistency, prior
165 experience in radiology could be assumed to affect repeatability, as supported by Verhoeven et al.
166 (2009).¹⁷ In our study, however, the intraobserver repeatability and pairwise analyses were almost
167 similar between all observers, regardless of experience. This may be a result of the observers having
168 the opportunity to get accustomed to the grading system before applying it. As subjectivity can
169 never be completely eliminated, there is always some inherent fluctuation within and between
170 individuals, leading to a higher likelihood of disagreement with increased grading repetitions.
171 Accordingly, we observed a slightly lower repeatability when all three grades were required to be
172 identical, compared to when only two were required.

173

174 When the images were grouped by breed, the agreement was lower in the Glen of Imaal terrier in
175 the intraobserver agreement analyses in two of the observers (A and B). The agreement was also
176 lower for the Glen of Imaal terrier in the breed-specific interobserver analysis. It is possible that this
177 is due to the slightly larger size – and thus longer limbs – of the Glen of Imaal terrier. Specifically,
178 as the x-ray beam is centered to the mid-radius, the longer antebrachium requires the beam to be
179 positioned farther away from the joint. As this study aimed to assess the repeatability of a
180 previously established imaging protocol, we did not make changes to the positioning of the patient
181 or the radiographic beam regarding the protocol described in the earlier study by Lappalainen et al.
182 2016. However, the effect of beam centering for recognizing elbow incongruity has been studied by
183 Murphy et al. (1998), who found that a congruent joint will not appear incongruent even if the
184 radiographic beam is not centered directly on the elbow.¹⁸ This was further supported in a later
185 study by Blond et al. (2005), who showed that only a proximal displacement affected the
186 recognition of normal elbows – a beam misalignment as large as 3 cm away from the center of the
187 joint did not affect the recognition of incongruity.¹⁰ Thus, it is unlikely that the larger size of the
188 Glen of Imaal terrier would explain the lower repeatability in this breed, but can be considered a
189 limitation of the imaging protocol for non-chondrodystrophic dog breeds. However, one breed-
190 related factor that could explain the noted difference in repeatability might be the amount of angular
191 deformity in the forelimb, which is likely to vary between breeds. The deformity may make it
192 difficult to position the joint in such a way that the medial and lateral border of the ulnar joint
193 surface would overlap perfectly in the image. The resulting double silhouette could have an effect
194 on measuring the joint space. Additionally, if the joint space measures very close to the boundary of
195 two grades, very small variations (such as a tenth of a millimeter) between the measurements may
196 result in a different INC grade. Subjectively, this could be the case especially between grades 0 and
197 1. The breeds selected for our study may represent different frequencies of these borderline cases of
198 the INC grades.

199

200 Although this study highlighted the importance of repeatability in INC grading, it is not without
201 limitations. For example, the frequencies of the different INC grades were not distributed equally.
202 The most dominant grade was INC1, comprising over half of the samples. Subjectively, among the
203 cases where demarcation between grades may have caused uncertainty, the most common were
204 borderline cases between INC0 and INC1. In the more severely affected cases, the joint space might
205 be easier to measure even in the presence of other deformities, especially when the joint space
206 would be wider than 3 mm. Therefore, it could be argued that if the study population would have
207 included a larger proportion of more severe cases (i.e. INC2 and INC3), the repeatability could
208 possibly be higher than what we report in the current study.

209

210 The INC grading as such does not take into account possible signs of osteoarthritis, which should
211 obviously be added to the protocol if used for screening purposes. Actually, the screening protocol
212 of the Finnish Kennel Club has incorporated osteoarthritis into the protocol, and a dog with signs of
213 degenerative joint disease cannot get INC grades 0 or 1.

214

215 In conclusion, our study showed that the radiographic grading of elbow incongruity by the INC
216 grading system had good intra- and interobserver repeatability in chondrodystrophic breeds such as
217 the dachshund, Glen of Imaal terrier and Skye terrier. This would make the INC grading system a
218 suitable method for screening elbow incongruity in these breeds. However, as this study only aims
219 to assess repeatability of the grading system, any recommendations on which grades should be
220 considered acceptable for each breed would require further investigations.

221

222 **List of author contributions**

223 Category 1

224 (a) Conception and Design: Pulkkinen HSM, Hyytiäinen HK, Laitinen-Vapaavuori OM,
225 Lappalainen AK

226 (b) Acquisition of Data: Pulkkinen HSM, Reunanen VLJ, Hyytiäinen HK, Lappalainen AK

227 (c) Analysis and Interpretation of Data: Pulkkinen HSM, Reunanen VLJ, Hyytiäinen HK, Junnila
228 JTT, Lappalainen AK

229

230 Category 2

231 (a) Drafting the Article: Pulkkinen HSM, Hyytiäinen HK, Laitinen-Vapaavuori OM, Lappalainen
232 AK

233 (b) Revising the Article for Intellectual Content: Pulkkinen HSM, Reunanen VLJ, Hyytiäinen HK,
234 Junnila JTT, Laitinen-Vapaavuori OM, Lappalainen AK

235

236 Category 3

237 (a) Final Approval of the Completed Article: Pulkkinen HSM, Reunanen VLJ, Hyytiäinen HK,
238 Junnila JTT, Laitinen-Vapaavuori OM, Lappalainen AK

239

240 **Acknowledgements**

241 We wish to thank research assistant Laura Parikka for her technical assistance in anonymization and
242 randomization of the radiographic data.

243

244 **References**

245 1. Parker HG, VonHoldt BM, Quignon P, Margulies EH, Shao S, Mosher DS, Spady TC, Elkahouloun
246 A, Cargill M, Jones PG, Maslen CL, Acland GL, Sutter NB, Kuroki K, Bustamante CD, Wayne
247 RK, Ostrander EA. An expressed *fgf4* retrogene is associated with breed-defining chondrodysplasia
248 in domestic dogs. *Science* 2009;325:995–998.

- 249
- 250 2. Lappalainen AK, Hyvärinen T, Junnila J, Laitinen-Vapaavuori O. Radiographic evaluation of
251 elbow incongruity in Skye terriers. *J Small Anim Pract* 2016;57(2):96–99.
- 252
- 253 3. Lau R. Inherited premature closure of distal ulnar physis. *J Am Ani Hosp Assoc* 1977;13(5):609–
254 612.
- 255
- 256 4. Wind AP, Packard ME. Elbow incongruity and developmental elbow diseases in the dog: Part II.
257 *J Am Ani Hosp Assoc* 1986; 22(6):725–730.
- 258
- 259 5. Ramadan RO, Vaughan LC. Premature closure of the distal ulnar growth plate in dogs--a review
260 of 58 cases. *J Small Anim Pract* 1978;19(11):647–667.
- 261
- 262 6. Knapp JL, Tomlinson JL, Fox DB. Classification of Angular Limb Deformities Affecting the
263 Canine Radius and Ulna Using the Center of Rotation of Angulation Method. *Vet*
264 *Surg* 2016;45(3):295–302.
- 265
- 266 7. Theyse LFH, Voorhout G, Hazewinkel HAW. Prognostic factors in treating antebrachial growth
267 deformities with a lengthening procedure using a circular external skeletal fixation system in dogs.
268 *Vet Surg* 2005;34:424–435.
- 269
- 270 8. IEWG. Explanation of grading according to IEWG and discussion of cases. Proceedings; 31st
271 Annual Meeting of the International Elbow Working Group. 2017.
- 272
- 273 9. OFA. Examining Elbow Dysplasia. Orthopedic Foundation for Animals; [accessed 31.10.2018].

274 <https://www.ofa.org/diseases/elbow-dysplasia>.

275

276 10. Blond L, Dupuis J, Beauregard G, Breton L, Moreau M. Sensitivity and specificity of
277 radiographic detection of canine elbow incongruence in an in vitro model. *Vet Radiol Ultrasoun*
278 2005;46(3), 210–216.

279

280 11. Mason DR, Schulz KS, Samii VF, Fujita Y, Hornof WJ, Herrgesell EJ, Kass PH. Sensitivity of
281 radiographic evaluation of radio-ulnar incongruence in the dog in vitro. *Vet Surg* 2002;31(2):125–
282 132.

283

284 12. Bartlett JW, Frost C. Reliability, repeatability and reproducibility: analysis of measurement
285 errors in continuous variables. *Ultrasound Obstet Gynecol* 2008;31:466–475.

286

287 13. Viera A, Garrett J. Understanding interobserver agreement: The Kappa statistic. *Fam Med*
288 2005;37:360–3633.

289

290 14. Verhoeven G, Coopman F, Duchateau L, Saunders JH, van Russen B, van Bree H. Interobserver
291 agreement in the diagnosis of canine hip dysplasia using the standard ventrodorsal hip-extended
292 radiographic method. *JSAP* 2007;48:387–393.

293

294 15. Fortrie RR, Verhoeven G, Breckx B, Duchateau L, Janssens L, Samoy Y, Schreurs E, Saunders
295 J, van Bree H, Vanderkerckhove P, Coopman F. Intra- and interobserver agreement on radiographic
296 phenotype in the diagnosis of canine hip dysplasia. *Vet Surg* 2015;44:467–473.

297

- 298 16. Geissbühler U, Drazovic S, Lang J, Howard J. Interrater agreement in radiographic canine hip
299 dysplasia evaluation. *Vet Rec* 2017.
300
- 301 17. Verhoeven GEC, Coopman F, Duchateau L, Bosmans T, Van Ryssen B, van Bree H.
302 Interobserver agreement on the assess ability of standard ventrodorsal hip-extended radiographs and
303 its effect on agreement in the diagnosis of canine hip dysplasia and on routine FCI scoring. *Vet*
304 *Radiol Ultrasoun* 2009;50(3):259–263.
305
- 306 18. Murphy ST, Lewis DD, Shiroma JT, Neuwirth LA, Parker RB, Kubilis PS. Effect of
307 radiographic positioning on interpretation of cubital joint congruity in dogs. *Am J Vet Res*
308 1998;59(11):1351–1357.
309

310 Table 1. The incongruity-grading system used for scoring elbow incongruity (as described by
 311 Lappalainen et al. 2016).²

Grade	Definition
INC0 (normal)	Even and narrow joint space; the width of the humeroulnar joint space measures < 1 mm.
INC1 (mild)	The width of the humeroulnar joint space measures 1–2 mm.
INC2 (moderate)	The width of the humeroulnar joint space measures 2–3 mm.
INC3 (severe)	The width of the humeroulnar joint space measures > 3 mm.

312 *INC, incongruity grade.*

313 Table 2. Frequencies of incongruity grades, by observer and measurement time point, for 220 elbow
 314 joint images (110 dogs) in three chondrodystrophic dog breeds.

Grade	Observer	A			B			C		
		Measurement time point 1	2	3	1	2	3	1	2	3
INC0		51	32	26	66	60	64	25	22	34
INC1		138	157	166	121	125	120	132	156	135
INC2		25	25	23	26	28	29	54	38	45
INC3		6	5	5	7	7	7	9	4	6

315 *INC, incongruity grade.*

316 Table 3. Intra-observer proportion of agreement and Fleiss' Kappa of incongruity grades for 220
 317 images of elbow joints from 110 dogs representing three chondrodystrophic dog breeds.

Observer	PA	Lower 95 % CI	Upper 95 % CI	Kappa	Lower 95 % CI	Upper 95 % CI
Three ratings identical						
A	0.777	0.72	0.83	0.690	0.635	0.745
B	0.714	0.65	0.77	0.678	0.623	0.733
C	0.705	0.64	0.76	0.623	0.568	0.677
At least two ratings identical						
A	1.000	0.98	1.00			
B	1.000	0.98	1.00			
C	0.991	0.97	1.00			

318 *INC, incongruity grade; PA, proportion of agreement; CI, confidence interval.*

319 Table 4. Interobserver proportion of agreement and weighted Kappa of incongruity grades for 220
 320 images of elbow joints from 110 dogs representing three chondrodystrophic dog breeds.

Agreement in question	PA	Lower 95 % CI	Upper 95 % CI	Kappa	Lower 95 % CI	Upper 95 % CI
All three agree	0.609	0.54	0.67	0.502*	0.448	0.557
At least two agree	0.991	0.97	1.00			
A vs. C	0.759	0.70	0.81	0.583	0.481	0.686
A vs. B	0.755	0.69	0.81	0.617	0.526	0.708
B vs. C	0.695	0.63	0.76	0.568	0.477	0.660

321 *Fleiss' Kappa; INC, incongruity grade; PA, proportion of agreement; CI, confidence interval.

322 Table 5. Intra-observer proportion of agreement and Fleiss' Kappa of incongruity grades for three
 323 identical gradings of 220 images of elbow joints from 110 dogs representing three
 324 chondrodystrophic dog breeds, analyzed by breed.

Observer	Breed	PA	Lower 95 % CI	Upper 95 % CI	Kappa	Lower 95 % CI	Upper 95 % CI
A							
	Glen of Imaal terrier	0.745	0.65	0.82	0.614	0.515	0.714
	Dachshund	0.850	0.73	0.93	0.727	0.619	0.835
	Skye terrier	0.759	0.62	0.87	0.727	0.611	0.844
B							
	Glen of Imaal terrier	0.651	0.55	0.74	0.562	0.463	0.662
	Dachshund	0.800	0.68	0.89	0.763	0.657	0.868
	Skye terrier	0.741	0.60	0.85	0.720	0.613	0.828
C							
	Glen of Imaal terrier	0.689	0.59	0.78	0.570	0.486	0.653
	Dachshund	0.767	0.64	0.87	0.663	0.555	0.772
	Skye terrier	0.667	0.53	0.79	0.595	0.464	0.726

325 *PA, proportion of agreement; CI, confidence interval.*

326 Table 6. Interobserver proportion of agreement and Fleiss' Kappa of incongruity grades for three
 327 identical gradings for 220 images of elbow joints from 110 dogs representing three
 328 chondrodystrophic dog breeds, analyzed by breed.

Breed	PA	Lower 95 % CI	Upper 95 % CI	Kappa	Lower 95 % CI	Upper 95 % CI
Glen of Imaal terrier	0.557	0.46	0.65	0.400	0.306	0.495
Dachshund	0.717	0.59	0.83	0.570	0.465	0.675
Skye terrier	0.593	0.45	0.72	0.528	0.410	0.645

329 *INC, incongruity grade; PA, proportion of agreement; CI, confidence interval.*

330 Figure 1A. Mediolateral radiograph of the antebrachium showing an example of an elbow with high intra
331 and intertester agreement of incongruity grade showing a single measurement made with the imaging
332 software. All three observers graded the elbow as INC1 on each of the three gradings. The image was
333 acquired using computed radiography with an automatic exposure detector (FUJI imaging plates, FUJIFILM
334 reader, 44 kV; 4,0 mAs, S-values of 100-300).

335 Figure 1B. Zoomed and cropped version of the same radiograph as Fig. 1A, centered over the
336 humeroulnar joint space.

337 Figure 2A. Mediolateral radiograph of the antebrachium showing an example of an elbow with low
338 intra and intertester agreement of incongruity grade. The elbow was graded INC0 three times and
339 INC1 six times. The image was acquired using computed radiography with an automatic exposure
340 detector (FUJI imaging plates, FUJIFILM reader, 44 kV; 4,0 mAs, S-values of 100-300).

341 Figure 2B. Zoomed and cropped version of the same radiograph as Fig. 2A, centered over the
342 humeroulnar joint space.