

Accepted Manuscript

Title: Inter-observer agreement for clinical examinations of foot lesions of sheep

Author: C.J. Phythian, P.C. Cripps, D. Grove-White, E. Michalopoulou, J.S. Duncan

PII: S1090-0233(16)30119-8

DOI: <http://dx.doi.org/doi: 10.1016/j.tvjl.2016.08.005>

Reference: YTVJL 4865

To appear in: *The Veterinary Journal*

Accepted date: 12-8-2016



Please cite this article as: C.J. Phythian, P.C. Cripps, D. Grove-White, E. Michalopoulou, J.S. Duncan, Inter-observer agreement for clinical examinations of foot lesions of sheep, *The Veterinary Journal* (2016), <http://dx.doi.org/doi: 10.1016/j.tvjl.2016.08.005>.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1 **Inter-observer agreement for clinical examinations of foot lesions of sheep**

2

3 Phythian, C.J.^{ac*}, Cripps, P.C.^a Grove-White, D.^a Michalopoulou, E.^a and Duncan, J.S.^a

4

5 ^a *Department of Epidemiology, University of Liverpool, Institute of Global Health and*
6 *Infection, and Population Health, Leahurst, Neston, CH64 7TE, United Kingdom*7 ^c *NMBU Sandnes, Section for Small Ruminant Research, Department of Production Animal*
8 *Clinical Science, Høyland, 4325 Sandnes, Norway*

9

10 *Corresponding author: Tel.: +47 51603510

11 *Email address:* clare.phythian@nmbu.no (C.J.Phythian)

12

13 *Current address:* NMBU, Section for Small Ruminant Research, 4325 Sandnes, Norway

14 **Highlights**

- 15 • Diagnostic evaluation of clinical examinations of ovine foot lesions.
- 16 • Inter-observer agreement and percentage disagreement were assessed.
- 17 • Some scoring disagreement occurred over the diagnosis of white line lesions.
- 18 • Trained observers can reliably score most common foot lesions of sheep.

19

20

Abstract

21 In sheep, the diagnosis of foot lesions is routinely based on physical examination of the
22 hoof. Correct diagnosis is important for the effective treatment, prevention and control of both
23 infectious and non-infectious causes of lameness. Therefore, the aim of this study was to
24 evaluate the level of inter-observer agreement for clinical examination of ovine foot lesions.
25 Eight observers of varying experience, training and occupation performed foot examinations on
26 a total of 1158 sheep from 38 farms across North England and Wales. On each farm, a group of
27 two to four observers independently examined a sample of 24 to 30 sheep to diagnose the
28 presence or absence of specific foot lesions including white line lesions (WL), contagious ovine

29 digital dermatitis (CODD), footrot (FR), inter-digital dermatitis (ID) and toe granuloma (TG).
30 The inter-observer agreement of foot lesion assessments was examined using Fleiss kappa (κ),
31 and Cohen's κ examined the paired agreement between the test standard observer (TSO) and
32 each observer.

33 Scoring differences with the TSO were examined as the percentage of scoring errors and
34 assessed for evidence of systematic scoring bias. With the exception of WL (maximum error
35 rate 33.3%), few scoring differences with the TSO occurred (maximum error rate 3.3%). This
36 suggests that observers can achieve good levels of reliability when diagnosing most of the
37 commonly observed foot conditions associated with lameness in sheep.

38

39 *Keywords:* Foot lesions; Sheep; Clinical diagnosis; Observer agreement.

40

41 **Introduction**

42 Lameness is a significant and serious global issue for sheep because of the pain,
43 discomfort and debilitation caused (Welsh et al., 1993; Ley et al., 1995; Fitzpatrick et al.,
44 2006). Research has identified that, globally, footrot is the most common cause of lameness in
45 sheep (Egerton et al., 1989; König et al., 2011). Consequently, a variety of strategies for control
46 and elimination of footrot have been devised. These include control approaches based on the
47 administration of systemic antibiotic treatments and culling of persistently-infected cases
48 (Wassink et al., 2010), and elimination strategies based on prophylactic vaccination and whole-
49 flock culling programs (Egerton et al., 2002; Egerton et al., 2004; Gurung et al., 2006) .

50

51 Whilst footrot may be a common cause of lameness (Kaler and Green, 2008a), clearly not
52 all lameness in sheep can be attributed to the condition. Contagious ovine digital dermatitis
53 (CODD), which results in severe lameness and loss of the hoof capsule, currently presents a
54 serious welfare concern for sheep in the UK (Winter, 2008). To date, there is limited
55 knowledge on the epidemiology of this disease and by comparison to footrot only a few recent
56 trials have examined the efficacy of systemic treatments (Duncan et al., 2011, 2012). In
57 addition, there are a number of other foot conditions, including separation and impaction of the
58 white line of the hoof, toe granulomas, interdigital-hyperplasia, septic- and osteo-arthritis,
59 which can also result in gait abnormalities of sheep (Winter, 2004; Winter and Arsenos, 2009;
60 Hodgkinson, 2011). Whilst infectious foot lesions remain the most important concern for flock
61 welfare, it has been suggested that these other hoof lesions, such as separation and impaction of
62 the white line (also known colloquially as ‘shelly hoof’), are underreported due to mis-
63 diagnosis and confusion with footrot cases (Conington et al., 2010a). This is of great
64 importance since the treatment and control points that are deemed to be effective for one foot
65 condition may not be relevant or appropriate for the control of another lesion or infectious
66 cause. The correct identification of a lesion or disease is essential not only for both animal
67 welfare reasons but also economic considerations in order to assess both the scale and
68 economic impact of the disease. Hence, the ability to correctly diagnose foot lesions is vital for
69 implementing prompt and effective treatments and the long-term prevention and control of
70 lameness in sheep flocks (Kaler and Green, 2008a; Kaler and Green 2008b).

71

72 The ease and accuracy of using diagnoses based on the clinical appearance of lesions
73 needs to be further considered given that there is considerable variation in the visual
74 appearance of ovine foot lesions (Kaler and Green, 2008a). Furthermore, there are recognised
75 differences in the interpretation and assessment of different foot lesions amongst differing

76 assessors, such as veterinary surgeons, farmers and researchers (Kaler and Green, 2008b).
77 Microbiological culture (Pitman et al., 1994) and PCR-based techniques (Moore et al., 2005;
78 Frosth et al., 2012) can be employed to complement clinical examination in the diagnosis of
79 some hoof pathologies. However, the time and financial cost of such methods preclude their
80 routine use. Thus, clinical examination by the producer or a veterinary surgeon remains the
81 mainstay for diagnosis of foot conditions in sheep. Consequently, the practical experience and
82 training of farm professionals and veterinarians in the recognition and correct diagnosis of
83 common foot lesions of sheep is an area that warrants further attention.

84

85 The diagnostic abilities of different observers can be examined in terms of the level of
86 inter-observer agreement or reliability. The reliability of both binary and categorical scoring
87 measures can be evaluated using agreement analysis methods such as the kappa coefficient (κ)
88 (Kaler et al., 2009). The agreement analysis presents the degree of observed agreement
89 compared to the agreement expected by chance (Feinstein and Cicchetti, 1990) and has been
90 widely used in veterinary research applications, for example to assess the observer reliability
91 for equine health and welfare indicator assessments (Burn et al., 2009) or lameness scores of
92 sheep (Kaler et al., 2009). The type of κ selected depends on the number of observers involved.
93 Fleiss's κ determines the reliability of multiple observers ($n > 2$) (Fleiss, 1981), whereas
94 Cohen's κ (Cohen, 1960) examines the reliability of paired assessments ($n = 2$) such as the
95 level of agreement between a study observer and a reference observer, such as the trainer (Burn
96 et al., 2009). κ can also be used to assess the level of agreement between each study observer
97 and a reference observer, such as the trainer (Burn et al., 2009). Several categorical systems for
98 scoring ovine hoof health conditions, and specifically footrot, have been developed and tested
99 (Egerton and Roberts, 1971; Raadsma et al., 1994; Conington et al., 2008; Foddai et al., 2012).
100 However, for routine on-farm assessments as conducted by producers and veterinarians it may

101 not be necessary to use such detailed scoring systems since a binary scale (presence or absence)
102 could provide sufficient information.

103

104 The objective of this study was to examine the level of inter-observer agreement for
105 specific ovine foot lesion conditions, using κ agreement analysis statistics and percentage error
106 rate results.

107

108 **Materials and methods**

109 *Study population*

110 The investigation was a cross-sectional study in which 38 farms, located in a 120 mile
111 radius of the University of Liverpool, School of Veterinary Science, Leahurst were recruited
112 through contact with their local veterinary practice. Once the informed consent of farmers was
113 obtained, each farm was requested to gather a sample of approximately 100 sheep for
114 assessment during July to November 2008. On the day of assessment, each sheep was then
115 assigned a numeric identifier in the order they entered the assessment pen and on each farm 30
116 sheep were selected for examination using a pre-determined random number system.

117

118 *Observer population*

119 A pool of eight observers was recruited from the University of Liverpool, School of
120 Veterinary Science comprising undergraduate veterinary and animal science students ($n = 3$)
121 and veterinary surgeons ($n = 5$). Observers were classified as ‘experienced’ if they had
122 undertaken clinical examinations and foot lesion diagnosis of sheep in the previous year (Table
123 1), those that did not meet these criteria were classified as inexperienced. On the basis of their
124 experience and role in the design and conduct of the study, observer 1 was designated the ‘test
125 standard observer’ (TSO) and used as the reference test for comparison. All observers were

126 provided with a scoring definition for each lesion, which they were requested to familiarise
127 themselves with together with example images of the specific lesions. In addition, observers
128 classed as ‘trained’ ($n = 5$) attended a one-day on-farm training session at the University of
129 Liverpool sheep farm in the diagnosis of foot lesions in sheep. The TSO performed assessments
130 on all study farms and was accompanied at each assessment visit by one to two observers who
131 performed independent clinical examinations of the same sheep on the same day.

132

133 *Foot examination*

134 Each observer independently performed a clinical examination of each foot of all
135 sample animals as described by Hodgkinson (2010). The absence or presence of any foot
136 lesion in each sheep was recorded. The following specific diagnoses were made based on the
137 descriptions of Winter (2004): white line lesion (WL) - separation and detachment of the
138 white line (‘shelly hoof’) with impaction or infection present: inter-digital dermatitis (ID) - a
139 raw to white, moist hairless area, progressing to inflammation, infection and necrosis of the
140 inter-digital skin: footrot (FR) - separation of the horn of the hoof, beginning at the junction
141 of the skin and horn near the heel, through to invasion of the sole with separation of
142 insensitive and sensitive laminae: contagious ovine digital dermatitis (CODD) - ulceration
143 around the coronary band, with or without loosening of the claw through to the total loss of
144 the hoof capsule and presence of a raw stump of sensitive laminae: toe granuloma (TG) -
145 strawberry-like growth of proud flesh, which may be covered with loose horn: inter-digital
146 hyperplasia (IH) - folds or protrusions of the skin of variable size located within the inter-
147 digital cleft, and pedal joint sepsis (PJS) - presence of heat, swelling and hair loss above the
148 coronary band, with or without discharging tracts of pus above the coronary band or
149 interdigital cleft. No diagnosis was recorded if it was not possible to make a specific
150 diagnosis based solely on the visual appearance of the foot. Each observer manually recorded

151 their findings on pre-tested recording charts. Observers were not provided with any clinical or
152 production information before each visit. During the visit, each study observer performed an
153 independent clinical examination and observers did not disclose or discuss their foot scores at
154 any stage. The study was approved by the University of Liverpool Ethics Committee
155 (RETH000287).

156

157 *Data analysis*

158 Data was analysed using Minitab version 16 and Stata version 13 (StataCorp LP). The
159 prevalence (percentage) and 95 percent confidence interval (95% CI) of each foot condition
160 was determined as the total number of sheep observed by the TSO with each foot condition
161 divided by the total number of sheep assessed.

162 The overall level of inter-observer reliability of multiple observer assessments ($n \geq 2$)
163 was determined by Fleiss's κ (Fleiss et al., 2003). As Fleiss's κ analysis provides an overall
164 agreement value and does not take account of observer characteristics i.e. 'experienced' versus
165 'inexperienced' assessors, the paired agreement between the TSO and each observer was
166 estimated using Cohen's κ statistic (Cohen, 1960). All κ results were interpreted according to
167 Fleiss et al., (2003), whereby values ≥ 0.75 suggested 'excellent', κ 0.40 - 0.75 indicated 'fair-
168 good', and $\kappa \leq 0.40$ suggested 'poor' levels of agreement.

169

170 As the κ analytical approach cannot identify whether systematic scoring differences
171 occur between pairs or groups of multiple observers, additional approaches were employed to
172 assess the level of observer disagreement in terms of scoring divergence from the TSO. Firstly,
173 scoring differences between the TSO and each observer (TSO score minus observer score)
174 were graphically represented and visually examined for evidence of systematic scoring bias i.e.
175 if an observer consistently scored one unit higher or lower than the TSO. For each observer, the

176 total number of lesions diagnosed by the TSO during paired assessments was calculated and the
177 number of paired scoring differences with the TSO divided by the total number of sheep
178 examined was expressed as a percentage (percentage error rate). Secondly, the proportion of
179 scoring differences with the TSO on each farm visit was plotted to assess if there was any effect
180 of increasing experience of foot assessments on the amount of scoring disagreements.
181 Observers were not provided with any clinical or production information before each visit.
182 During the visit, each study observer performed an independent clinical examination and
183 observers did not disclose or discuss scores at any stage.

184

185 **Results**

186 A total of 4632 feet from 1158 sheep were examined for the presence of specific foot
187 lesions. From the pool of eight observers, a varying group of two to three observers, including
188 the TSO, independently examined the feet of 24 to 30 animals on each farm. Data recorded by
189 the TSO indicated that over half of the population ($n = 610$, 52.6%) was observed to have at
190 least one recorded condition in one or more feet. The most frequently observed lesion was WL
191 (49.1%) and few cases of FR, ID, TG and CODD were recorded (Table 2). No cases of pedal
192 joint sepsis cases were identified and there were insufficient observations of interdigital
193 hyperplasia recorded by observers 2 – 8 to permit evaluation of the reliability, error rates for
194 these foot conditions.

195

196 Overall level of inter-observer reliability was interpreted as ‘excellent’ for assessments of
197 CODD (κ 0.72, 95% CI 0.71 - 0.77) being ‘fair-good’ for WL (κ 0.47, 95% CI 0.46 - 0.47) and
198 TG (κ 0.65 95% CI 0.46 - 0.85). Fleiss κ values for FR (κ 0.49, 0.35 - 0.63) and ID (κ 0.49, 95%
199 CI 0.37 - 0.65) diagnoses were ‘fair to good’ but the lower 95% confidence intervals for these
200 lesions indicated some ‘poor’ levels of κ agreement occurred (Table 3).

201

202 With the exception of WL assessments, the majority of paired assessments with the TSO
203 showed a low level ($\leq 3.3\%$ error rate) of scoring disagreement (Table 4). Graphical
204 representation of the frequency of scoring differences for the diagnosis of white line lesions
205 suggested there were some systematic scoring differences in the diagnosis of WL by several
206 observers (Fig. 1). The evaluation of the effect of time on agreement with the TSO was limited
207 to three observers (observers 3, 4, 7) who each performed ten or more study visits. Other study
208 observers performed insufficient visits to facilitate this evaluation. Graphical representation of
209 the proportion of scoring differences between the TSO and observers 3, 4 and 7 suggested there
210 was no effect of increasing number of farm visits on the level of scoring disagreements (Fig. 2).

211

212 **Discussion**

213 The objective of this study was to evaluate the performance of eight assessors of varying
214 experience, training and occupation on their ability to agree on the diagnosis of a range of
215 specific foot conditions of sheep based solely on clinical examination and by using a binary
216 scoring system. For the purposes of this study, farmers provided a group of sheep for
217 assessment from which 30 animals were randomly selected for foot lesion examination.
218 Previous work by Foddai et al. (2012) found high levels of inter-observer reliability when three
219 observers used a combination of video, photographic images and post-mortem feet specimens
220 to score lesions and foot shape using an ordinal scoring system. Assessments of lesion images
221 produce higher levels of observer agreement compared to assessments of cadaver foot
222 specimens (Foddai et al., 2012) and may reflect the more controlled observational conditions
223 provided by image-based studies or the selection of lesion images that illustrate clear,
224 characteristic signs of specific disease. When comparing the reliability and diagnostic test
225 results of different studies, as well as considering the type of scoring system used it is also

226 important to evaluate the conditions for assessment. It is possible that the levels of observer
227 agreement that are attainable during on-farm studies may be with vastly different from those of
228 image-based studies given the less controlled observational conditions and the need for
229 handling of animals, which may introduce observational errors (Foddai et al., 2012). Therefore,
230 a key aspect of this study was to test diagnostic abilities under conditions that producers and
231 veterinary surgeons, who routinely conduct ovine foot examinations, are exposed to. Whilst the
232 reliability of footrot scoring systems has been previously examined (Conington et al., 2008;
233 Foddai et al., 2012), to the best of our knowledge, this is the first study to examine the inter-
234 observer agreement of eight observers who examined and scored a large number of feet ($n =$
235 4632) from sheep managed under differing farm production systems for eight specific lesions.

236

237 Overall, Fleiss kappa results indicate that acceptable levels of reliability were achieved
238 for the combined FR and ID scores, CODD and TG. Limitations in the availability of methods
239 of agreement analysis and issues with the use of κ for the evaluation of observer reliability are
240 well-recognised (Feinstein and Cicchetti, 1990; Burn et al., 2009; Foddai., 2012). The strong
241 influence of lesion prevalence on κ estimates can be a particular issue for reliability studies
242 conducted under field conditions (Burn et al., 2009). Therefore, these results should be viewed
243 in light of the low prevalence of certain foot conditions, since this can reduce the level of κ and
244 subsequent interpretation of the degree of inter-observer agreement achieved (Feinstein and
245 Cicchetti, 1990; Burn et al., 2009). Another limitation with κ is the inability to quantify the
246 level of scoring disagreement. In addition, whilst Fleiss's κ gives a useful indication of
247 agreement between multiple observers the method does not take account of any biases due to
248 observer characteristics, for example, 'experienced' vs. 'trained' vs. 'inexperienced' that may
249 arise in multiple observer combinations. Observer 1 was selected as a 'pseudo-gold standard'
250 reference standard using the approach of Burn et al., (2009) in order to compare paired inter-

251 observer agreement and scoring divergence with the trainer and to assess for evidence of
252 systematic scoring bias (Bland and Altman, 1986). Since clinical examinations are subjective,
253 in some cases, it is possible that some observer scoring divergence from their trainer could
254 represent a closer approximation to the true (latent) foot condition. For example, the paired κ
255 agreement with the TSO ranged from poor to good for assessment of ID and FR but few
256 scoring errors (maximum error rate 3.3%) were found for both of these lesions. Here, the κ
257 results for ID and FR are considered to reflect the low number of animals observed in the study
258 and this likely affected the cross-tabulation of results, required for agreement analysis. Kappa
259 values are generally provided on a scale of 0 to 1 but negative values do arise and indicate
260 poorer agreement than that expected by chance alone (Cohen, 1960). A negative κ value arose
261 in the 95% CI for the paired assessments of TSO and observer 7, which reflected the very low
262 number of animals that were observed with footrot during the paired scoring sessions.

263

264 Other studies have examined the diagnostic abilities of other assessors including farmers
265 and veterinarians. Direct comparison of reliability studies can be complicated by differences in
266 the scoring systems used, selection of material used for assessment; ranging from photographic
267 images, video clips, post-mortem specimens (Foddai et al., 2012) or live sheep (Conington et
268 al., 2010b), and the context or conditions for assessment. Earlier research into the diagnostic
269 abilities of sheep veterinarians and producers found that $\geq 94\%$ of veterinarians correctly
270 diagnosed ID, FR, CODD and TG (Kaler and Green, 2008a). By contrast, only 26% of sheep
271 farmers could correctly diagnose the same lesions (Kaler and Green, 2008a). Reliability results
272 from the present study appear to concur with results of Kaler and Green (2008a), suggesting
273 that when present these infectious foot lesions can be readily identified by experienced and/or
274 trained assessors.

275

276 The high prevalence of WL identified in this study is in contrast to other studies from
277 Australia (Egerton et al., 1989), England (Grogono-Thomas and Johnston, 1997) and Sweden
278 (König et al., 2011), which suggest that footrot is the most commonly diagnosed foot condition
279 of lambs and sheep. Indeed, according to farmer reports, footrot is the most prevalence foot
280 lesion identified in English sheep flocks with a reported within-flock prevalence of 10% (Kaler
281 and Green, 2008a). However, the WL prevalence findings of the present on-farm study (49%)
282 do concur with those of Conington et al. (2010b), who assessed foot health scores from 27
283 flocks across the UK. With the exception of the Texel breed, white line separation was the
284 lesion found at highest (40%) prevalence (Conington et al., 2010b). In the present study,
285 participating farms were a convenience sample selected according to farm type and consent
286 thus the presented results cannot be interpreted as prevalence estimates. However, these farms
287 were considered to be representative of commercial sheep farming systems in England and
288 Wales and these results may highlight some interesting regional trends in sheep managed in
289 these flocks. To the author's knowledge, these farms had not been involved in previous
290 research or training on sheep lameness. Although, it is possible that the low level of FR and ID
291 identified here may suggest that farms with good ovine footrot control programs were recruited.
292 These findings may also reflect the management, environmental and climatic conditions at the
293 time of assessment that resulted in few sheep being diagnosed with these infectious lesions. The
294 vast majority of WL lesions observed in the current study were restricted to separation of the
295 hoof without impaction and infection of the white line (Winter and Arsenos, 2009). These
296 observations are in agreement with a single-flock trial, which identified a high prevalence of
297 WL of relatively minor degree of separation and an absence of other foot lesions (Wheeler et
298 al., 2013). It is possible that many sheep have a mild degree of white line separation, which
299 may be considered clinically insignificant, or missed during routine foot inspections. Co-
300 existing minor WL lesions might also not be recorded during inspections focused on the

301 detection of other ovine foot lesions, which might explain the prevalence findings reported
302 here. In spite of a high proportion of white line lesions, there does not appear to be a strong
303 association with a high level of flock lameness (Phythian et al., 2013) as often occurs with
304 footrot (Kaler et al., 2011). The significance of seemingly minor WL separation on foot health
305 and sheep welfare is not fully understood. Whilst there is some genetic heritability to ovine
306 white line degeneration (Conington et al., 2010a), currently the prevention and control of this
307 condition, predisposition to other hoof diseases and subsequent flock lameness prevalence is
308 unknown.

309

310 Interestingly, the data revealed systematic scoring differences consistently arose over WL
311 assessments. The seemingly minor degree of separation of the white line and potential poor
312 differentiation of mild WL lesions might explain the level of scoring disagreement found here.
313 Although a large number of feet were examined during the course of the study, the varying
314 number of farm visits conducted by all observers limited the ability of the study to fully assess
315 the effect of training and experience on diagnostic performance and no conclusions can be
316 reached in this respect. There are some trends in the data to suggest that following on-farm
317 training, inexperienced and trained observers (observers 3, 4 and 7) did not become more
318 reliable in WL diagnoses over the course of examining more than 1000 feet. However, the
319 results are limited to observers 3, 4 and 7 since they were the only observers that undertook ten
320 or more farm visits, which facilitated the evaluation of reliability over time. Further evaluation
321 of the effect of experience gained over a longer period of time and assessing whether a re-
322 calibration session is beneficial for inexperienced observers would be valuable here.

323

324 In addition to observer experience and training, scoring errors can also arise due to
325 misclassification that may be associated with the type of scoring system used. Misclassification

326 of disease errors may have arisen here over the diagnosis of inter-digital dermatitis and footrot
327 due to splitting of the scoring system into two distinct categories. With footrot lesions the
328 different clinical outcomes that can arise are due in part to the strain of *Dichelobacter nodosus*
329 involved (Moore et al., 2005), host susceptibility and genetic resistance (Emery et al., 1984).
330 No infectious disease model of inter-digital dermatitis has yet been demonstrated, although
331 some consider ID to be a continuum of clinical signs of a single disease (virulent and benign
332 footrot) (Egerton and Roberts, 1971). In the present study, simple binary scoring scales were
333 used to score benign and virulent footrot separately. However, a simple presence and absence
334 binary scoring system may clearly not accurately describe the continuum of disease signs
335 observed in footrot cases. With further training, assessors could be trained to grade the severity
336 of these footrot lesions using more detailed and categorical footrot scoring systems, such as
337 those of Egerton and Roberts (1971), Raadsma et al. (1994), and Nieuwhof et al. (2008). This
338 may be desirable for examining the effectiveness of different treatments or disease elimination
339 program (Egerton et al., 2004).

340

341 Very few cases ($n = 3$) were recorded by the TSO with no diagnosis. These cases were
342 considered to represent developing and early lesions that could not be defined as a specific
343 condition based solely on the visual appearance of the foot. Therefore, in some instances,
344 microbiological and molecular biological testing (Moore et al., 2005; Frosth et al., 2012) may
345 be required to support clinical examinations. The high levels of inter-observer kappa agreement
346 achieved for CODD may be attributed to the clear scoring definition and training provided, or
347 the ease of recognising this foot condition in sheep. Further training of observers in a recently
348 developed categorical CODD scoring system (Angell et al., 2015) could facilitate clinical trials
349 and further research into this condition, which is recognised to be of increasing importance in
350 UK flocks and presents serious concerns for sheep welfare (Duncan et al., 2011, 2012). Despite

351 this, on-farm experiences suggest that outbreaks of CODD are frequently mis-diagnosed as
352 footrot by producers who are unaware of this condition and/or the different physical features of
353 the disease (personal observation, CJ Phythian). However, in this study observers who were
354 previously unaware and unfamiliar with this foot lesion became competent at diagnosing
355 CODD. Such findings could inform disease awareness campaigns and highlights the value and
356 role of sheep veterinarians in the prompt diagnosis, treatment and control of flock lameness.

357

358 **Conclusion**

359 FR, ID, CODD and TG were consistently diagnosed by observers (maximum error rate
360 3.3%) while WL, the lesion most commonly recorded in this study, was missed or
361 misdiagnosed by some observers (maximum error rate 33.3%). The consequences for
362 researchers and veterinary practitioners may be that in spite of training and experience a degree
363 of measurement error and scoring disagreement can occur when using clinical examinations to
364 diagnose common foot lesions in sheep. This may result in under- or over-reporting of
365 prevalence estimates of some foot lesions during field studies, which needs to be considered
366 when assessing the treatment, control and prevention of lameness in sheep to ensure that the
367 optimal plans and advice are targeted at the correct lesion(s). In addition to further training of
368 assessors, in some cases, diagnoses based on visual inspections of ovine feet may need to be
369 supplemented by other tools such as molecular diagnostic testing.

370

371 **Conflict of interest statement**

372 None of the authors has any financial or personal relationships that could
373 inappropriately influence or bias the content of the paper.

374

375 **Acknowledgements**

376 The authors gratefully acknowledge the funding provided by Defra as part of the
377 AW1025 project (Development of indicators for the on-farm assessment of sheep welfare),
378 the support of the expert panel, study farms and on-farm assistance of Phillip Jones, Daniel
379 Hughes, Stephen Beer and Rachel Wyllie. We also thank the two anonymous reviewers for
380 their constructive comments on the earlier versions of this manuscript. Preliminary results
381 were presented in part as an Abstract at the First European College of Small Ruminant Health
382 Management Conference, Athens, Greece, 29-30 October 2011.

383

384 **References**

385 Angell, J.W., Bludell, R., Grove-White, D. W., Duncan, J.S., 2015. Clinical and
386 radiographic features of contagious ovine digital dermatitis and a novel lesion grading system.
387 *Veterinary Record* doi:10.1136/vr.102978.

388 Bland, J.M., Altman, D.G., 1986. Statistical methods for assessing agreement between two
389 methods of clinical measurement. *Lancet* 1, 307-310.

390 Burn, C.C., Pritchard, J.C., Whay, H.R., 2009. Observer reliability for working equine
391 welfare assessment: problems with high prevalences of certain results. *Animal Welfare* 18,
392 177-187.

393 Cohen, J., 1960. A coefficient of agreement for nominal scales. *Educational and*
394 *Psychological, Measurement* 20, 37-46.

395 Conington, J., Hosie, B., Nieuwhof, G., Bishop, S., Bünger, L., 2008. Breeding for
396 resistance to footrot – the use of hoof lesion scoring to quantify footrot in sheep. *Veterinary*
397 *Research Communications* 32, 583-589.

398 Conington J., Nicoll, L., Mitchell, S., Bünger, L., 2010a. Characterisation of white line
399 degeneration in sheep and evidence for genetic influences on its occurrence. *Veterinary*
400 *Research Communications* 34, 481– 489.

401 Conington, J., Speijers, M.H.M., Carson, A., Johnston, S., Hanrahan, S., 2010b. Foot
402 health in sheep – prevalence of hoof lesions in UK and Irish sheep. In: *Proceedings of the*
403 *British Society of Animal Science Annual Conference*, Belfast, Northern Ireland.

404 Duncan, J.S., Grove-White, D., Phythian, C.J., Oultram, J., Van Dijk, J., Carter, S.D.,
405 Cripps, P.J., Williams, H., 2011., Effects of parenteral amoxicillin on recovery rates and new
406 infection rates for contagious ovine digital dermatitis in sheep. *Veterinary Record* 169, 606.

- 407 Duncan, J.S., Grove-White, D., Moks, E., Oultram, J.W., Phythian, C.J., Williams, H.W.,
408 2012. Impact of footrot vaccination and antibiotic therapy on footrot and contagious ovine
409 digital dermatitis: a split flock randomised treatment trial. *Veterinary Record* 70, 462.
- 410 Egerton, J.R., Roberts, D.S., 1971. Vaccination against ovine foot-rot. *Journal of*
411 *Comparative Pathology* 8, 179-185.
- 412 Egerton, J.R., Yong, W.K., Riffkin, G.G., 1989. Footrot and foot abscess in ruminants.
413 CRC Press Inc, Florida, USA.
- 414 Egerton, J.R., Ghimire, S.C., Dhungyel, O.P., Shrestha, H.K., Joshi, H.D., Joshi, B.R.,
415 Abbott, K.A., Kristo, C., 2002. Eradication of virulent footrot from sheep and goats in an
416 endemic area of Nepal and an evaluation of specific vaccination. *Veterinary Record* 151, 290-
417 295.
- 418 Egerton, J.R., Seaman, J.T., Walker, R.I., 2004. Eradication of virulent footrot from New
419 South Wales. In: *Proceedings of the 13th Intentional Symposium and 5th Conference on*
420 *Lameness in Ruminants*, 11-15th February 2004, Maribor, Slovenija, 215-221.
- 421 Emery, D.L, Stewart, D.J., Clark BL., 1984. The comparative susceptibility of five breeds
422 of sheep to foot-rot. *Australian Veterinary Journal* 61, 85-88.
- 423 Feinstein, A.R., Cicchetti, D.V., 1990. High agreement but low kappa .1. The problems of
424 2 paradoxes. *Journal of Clinical Epidemiology* 43, 543-549.
- 425 Fitzpatrick, J., Scott, M., Nolan, A., 2006. Assessment of pain and welfare in sheep. *Small*
426 *Ruminant Research* 62, 55-61.
- 427 Fleiss, J. L., Levin, B. and Paik, M. C. 2003. *The Measurement of Interrater Agreement*,
428 *In: Statistical Methods for Rates and Proportions*, Third Edition, John Wiley & Sons, Inc.,
429 Hoboken, New Jersey, USA.
- 430 Foddai, A., Green, L.E., Mason, S.A., Kaler, J., 2012. Evaluating observer agreement of
431 scoring systems for foot integrity and footrot lesions in sheep. *BMC Veterinary Research* 8,
432 65.
- 433 Frosth, S., Slette-meås, J.S., Jørgensen, H.J., Angen, O., Aspán A., 2012. Development
434 and comparison of a real-time PCR assay for detection of *Dichelobacter nodosus* with
435 culturing and conventional PCR: harmonisation between three laboratories. *Acta Veterinaria*
436 *Scandinavica* 54, 6.
- 437 Grogono-Thomas, R. & Johnston, A. M. 1997. A study of ovine lameness. MAFF Final
438 Report. MAFF open contract OC59 45K.
- 439 Gurung, R.B., Dhungyel, O.P., Tshering, P., Egerton, J.R., 2006. The use of an autogenous
440 *Dichelobacter nodosus* vaccine to eliminate clinical signs of virulent footrot in a sheep flock in
441 Bhutan *Veterinary Journal* 172, 356-363.

- 442 Hodgkinson, O., 2010. The importance of feet examination in sheep health management.
443 Small Ruminant Research 92, 67-71.
- 444 Kaler, J., Green, L.E., 2008a. Naming and recognition of six foot lesions of sheep using
445 written and pictorial information: a study of 809 English sheep farmers. Preventive Veterinary
446 Medicine 83, 52-64.
- 447 Kaler, J., Green, L.E., 2008b. Recognition of lameness and decisions to catch for
448 inspection among sheep farmers and specialists in GB. BMC Veterinary Research 4, 41.
- 449 Kaler, J., Wassink, G.J., Green, L.E., 2009. The inter- and intra-observer reliability of a
450 locomotion scoring scale for sheep. The Veterinary Journal 180, 189-194.
- 451 Kaler, J., George, T.R.N., Green, L.E. 2011. Why are sheep lame? Temporal associations
452 between severity of foot lesions and severity of lameness in 60 sheep. Animal Welfare 20, 433-
453 438.
- 454 König, U., Nyman, A.K., de Verdier K., 2011. Prevalence of footrot in Swedish slaughter
455 lambs. Acta Veterinaria Scandinavica. 53, 27.
- 456 Ley, S.J., Waterman, A.E., Livingston, A., 1995. A field study of the effect of lameness on
457 mechanical nociceptive thresholds in sheep. Veterinary Record 137, 85-87.
- 458 Moore, L.J., Wassink, G.J., Green, L.E., Grogono-Thomas, R., 2005. The detection and
459 characterisation of *Dichelobacter nodosus* from cases of ovine footrot in England and Wales
460 Veterinary Microbiology 108, 57-67.
- 461 Nieuwhof, G.J., Conington, J., Bünger, L., Haresign, W., Bishop, S.C., 2008. Genetic and
462 phenotypic aspects of foot lesion scores in sheep of different breeds and ages. Animal 2, 1289-
463 1296.
- 464 Pitman, D.R., Palmer, M.A., Depiazzi, L.J. 1994. The laboratory culture of *Dichelobacter*
465 *nodosus* in a footrot eradication program. Australian Veterinary Journal 71, 109-112.
- 466 Phythian, C.J., Cripps, P.C., Grove-White, D., Jones, P.H., Michalopoulou, E., Duncan,
467 J.S. 2013. Observing lame sheep: evaluating test agreement between group-level and
468 individual animal methods of assessment. Animal welfare 22, 417-422.
- 469 Raadsma, H.W., Egerton, J.R., Wood, D., Kristo, C., Nicholas, F.W., 1994. Disease
470 resistance in Merino sheep. III. Genetic variation in resistance to footrot following challenge
471 and subsequent vaccination with a homologous rDNA pilus vaccine under both induced and
472 natural conditions. Journal of Animal Breeding and Genetics 111, 367-390.
- 473 Wassink, G.J., King, E.M., Grogono-Thomas, R., Brown, J.C., Moore, L.J., Green L.E.,
474 2010. A within farm clinical trial to compare two treatments (parenteral antibacterials and
475 hoof trimming) for sheep lame with footrot. Preventive Veterinary Medicine 96. 93-103.

476 Welsh, E.M., Gettinby, G., Nolan, A.M., 1993. Comparison of a visual analogue scale and
477 a numerical rating scale for assessment of lameness, using sheep as a model. *American Journal*
478 *of Veterinary Research* 54, 976-983.

479 Winter, A., 2004. Lameness in sheep: 1. Diagnosis. *In Practice* 26, 58-63.

480 Winter, A.C., 2008. Lameness in sheep. *Small Ruminant Research* 76, 149-153.

481 Winter, A.C., Arsenos, G., 2009. Diagnosis of white line lesions in sheep. *In Practice* 31,
482 17–21.

483 Wheeler, K., Phythian, C.J., Phillips, K., 2013. Effect of oral biotin supplementation on
484 the observed prevalence and severity of white line lesions in a lowland flock. In: *Proceedings*
485 *of the 13th International Symposium and 9th Conference on Lameness in Ruminants*, 11-14th
486 August 2013, Bristol, UK.

Figure legends

487

488

489 Fig. 1. Frequency of scoring differences between the paired foot examinations of the test
490 standard observer (TSO) and each study observer (2 - 8) for diagnosis of white line (WL).

491 Fig. 2. Frequency of scoring differences between paired examinations of trained and
492 inexperienced observers 3, 4 and 7 with the test standard observer (TSO) for white line lesion
493 (WL) diagnosis.

Data is presented only for observers who conducted foot examinations on ten or more farms to permit evaluation of the effect of time on observer reliability.

Table 1. Description of the observer population

Observer	Training	Experience	Occupation
1	Trainer	Experienced	Veterinary surgeon
2	Trained	Inexperienced	Veterinary surgeon
3	Trained	Inexperienced	Animal science student
4	Trained	Inexperienced	Veterinary science student
5	Untrained	Inexperienced	Veterinary surgeon
6	Untrained	Experienced	Veterinary surgeon
7	Trained	Inexperienced	Animal science student
8	Trained	Experienced	Veterinary surgeon

Table 2. Total number and percentage of sheep ($n = 1158$) observed with each foot lesion by the Test Standard Observer (TSO).

Diagnosis	Total n observed	Percentage (%) observed (95% CI)
White line lesion	569	49.1 (46.3 – 52.0)
Inter-digital dermatitis (ID)	11	0.9 (0.4 – 1.5)
Footrot	14	1.0 (0.6 – 1.8)
Contagious ovine digital dermatitis	16	1.4 (0.7 – 2.1)
Toe granuloma	16	1.4 (0.7 – 2.1)
Interdigital hyperplasia	5	0.4 (0.1 – 0.8)
No diagnosis	3	0.3 (0 – 0.6)

Table 3. Inter-observer agreement (Fleiss's kappa, 95% confidence interval), and paired agreement between the test standard observer and observers 2 – 8 (Cohen's kappa, 95% confidence interval) for diagnoses of specific ovine foot lesions

Diagnosis	Fleiss κ (95% CI)	Cohen's κ (95 % CI) by observer						
		2	3	4	5	6	7	8
WL	0.47 (0.46 - 0.47)	0.63 (0.35 - 0.90)	0.42 (0.36 - 0.48)	0.46 (0.40 - 0.52)	0.28 (0.04 - 0.52)	0.34 (0.02 - 0.67)	0.53 (0.43 - 0.64)	0.70 (0.52 - 0.88)
ID	0.49 (0.35 - 0.63)	a	0.25 (0.02 - 0.49)	0.73 (0.51 - 0.96)	0.78 (0.49 - 1)	a	0.67 (0.05 - 1)	a
FR	0.49 (0.37 - 0.65)	a	0.58 (0.34 - 0.82)	0.55 (0.27 - 0.83)	a	a	0.40 (-0.15 - 0.94)	a
CODD	0.72 (0.71 - 0.77)	1 (0.99 - 1)	0.55 (0.27 - 0.83)	0.75 (0.51 - 0.99)	1 (0.99 - 1)	a	0.68 (0.47 - 0.89)	1 (0.99 - 1)
TG	0.65 (0.46 - 0.86)	a	0.71 (0.44 - 0.98)	0.57 (0.26 - 0.88)	0.65 (0.20 - 1)	a	0.56 (0.25 - 0.88)	1 (0.99 - 1)

^a Insufficient observations of foot condition for kappa analysis

WL, white line lesion; ID, inter-digital dermatitis; FR, footrot; CODD, contagious ovine digital dermatitis; TG, toe granuloma

Table 4. Observer error rate expressed as the percentage (%) of scoring differences between the paired examinations with the test standard observer (TSO) and the total number (*n*) of lesions diagnosed by the TSO for each of the observer paired examinations

		Observer						
		2	3	4	5	6	7	8
Total <i>n</i> sheep examined		86	907	771	60	30	270	60
WL	percentage error	6.9	28.8	27.4	31.7	33.3	21.9	15.0
	<i>n</i> by TSO	9	391	372	37	12	175	9
ID	percentage error	1.2	1.9	0.7	3.3	0	0.4	0
	<i>n</i> by TSO	1	7	9	4	0	1	0
FR	percentage error	1.2	1.1	1.0	1.7	3.3	1.1	0
	<i>n</i> by TSO	1	11	10	0	0	3	0
CODD	percentage error	0	0.9	0.5	0	0	2.9	0
	<i>n</i> by TSO	0	7	6	1	0	9	3
TG	percentage error	1.1	0.3	0.4	0	0	0	0
	<i>n</i> by TSO	2	3	5	1	0	2	1

WL, white line lesion; ID, inter-digital dermatitis; FR, footrot; CODD, contagious ovine digital dermatitis; TG, toe granuloma.