© 2013 Universities Federation for Animal Welfare
The Old School, Brewhouse Hill, Wheathampstead,
Hertfordshire AL4 8AN, UK
www.ufaw.org.uk

Animal Welfare 2013, 22: 417-422
ISSN 0962-7286
doi: 10.7120/09627286.22.4.417

Observing lame sheep: evaluating test agreement between group-level and individual animal methods of assessment

CJ Phythian*[†‡], PC Cripps[§], D Grove-White[§], PH Jones[‡], E Michalopoulou[§] and JS Duncan[§]

[†] Animal Welfare and Behaviour Group, School of Veterinary Science, University of Bristol, Langford, Bristol BS40 5DU, UK
[‡] Department of Epidemiology, University of Liverpool, Institute of Global Health and Infection, and Population Health, Leahurst, Neston CH64 7TE, UK
[§] School of Veterinary Science, University of Liverpool, Leahurst, Neston CH64 7TE, UK
* Contact for correspondence and requests for reprints: C.J.Phythian@bristol.ac.uk

Abstract

For on-farm sheep welfare assessment, a reliable, simple and robust method is required to assess the level of flock lameness. This study examined the level of test agreement for two binary lameness scoring systems for sheep. The first was a group-level lameness assessment of sheep performed on ungathered sheep at pasture and was termed group observation method (GOM). The second method of lameness assessment was performed after gathering of the sheep and involved close observation of the gait of individual sheep in a handling pen and was termed individual animal gait assessment (IAGA). Following individual gait assessment, each sheep was also examined for the presence of specific foot and limb lesions: white line lesions (WL); inter-digital dermatitis (ID); footrot (FR); contagious digital dermatitis (CODD); toe granuloma (TG); and joint swellings (JS). A total of 3,074 sheep were assessed from 40 flocks in North England and Wales by one assessor. Test agreement between the assessment methods was found to be good as judged by linear regression and Bland-Altman plots. The method of group observation identified a slightly higher proportion of lame sheep compared to the individual animal examination and also appeared to be a more feasible on-farm method of observation. Over half of the sample sheep were identified with WL but this did not appear to be associated with a high level of lameness (as assessed by IAGA) with just under 12% of sheep with WL being identified as lame. In contrast, the percentage of lame sheep was most closely associated with CODD and over 80% of animals with this lesion were scored as lame.

Keywords: *animal welfare, foot lesion, lameness, on-farm assessment, sheep, test agreement*

Introduction

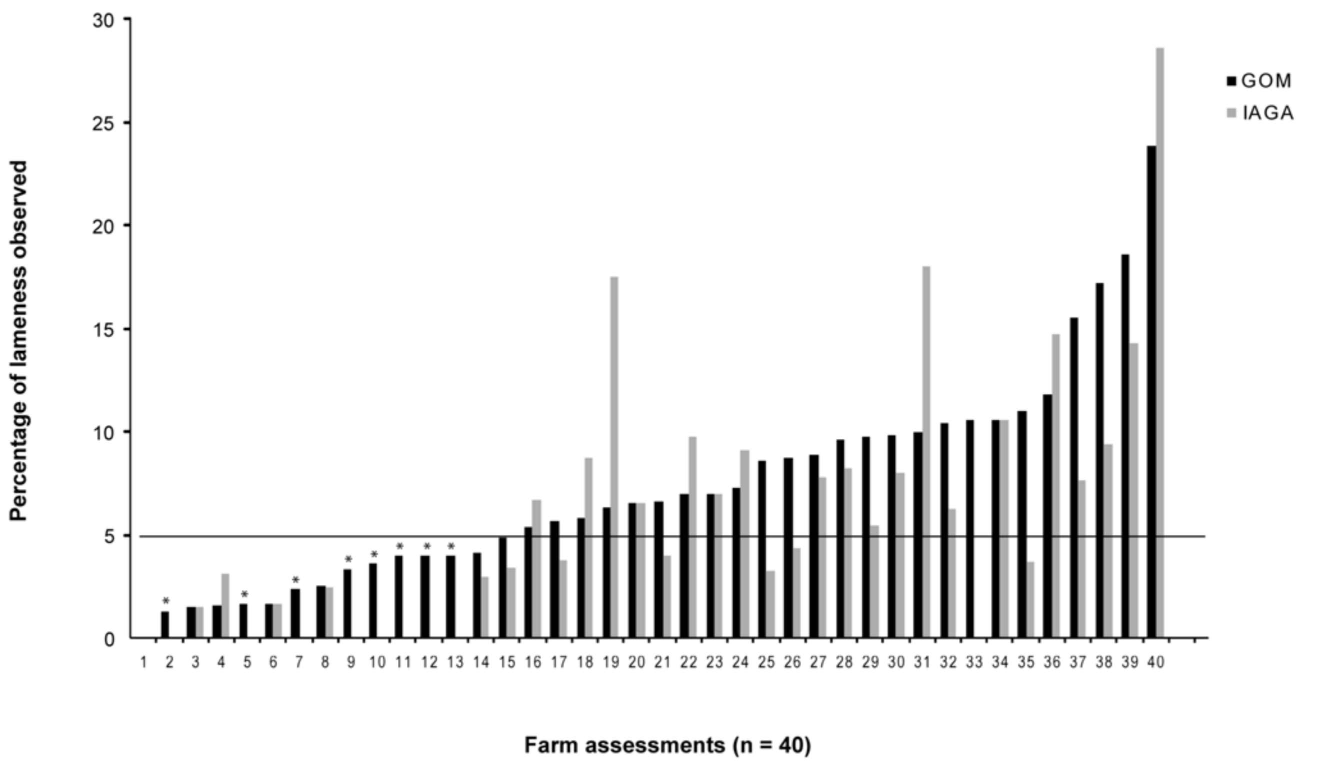
Lameness in sheep is an extremely important welfare and economic issue for the sheep industry (Nieuwhof & Bishop 2005; FAWC 2011). In order to inform welfare assessments and intervention strategies, farmers, veterinarians and welfare inspectors require valid and feasible means of measuring lameness in sheep flocks. The level of lameness within the flock is often assessed by farmers and others by observing the behaviour of groups of sheep at pasture. Alternatively, sheep are collected and handled to facilitate a closer inspection of gait and/or a physical foot examination of individual animals (Hodgkinson 2010). There can be differences in the interpretation and assessment of sheep lameness by veterinary surgeons and farmers (Harkins 2005; Kaler & Green 2008a,b), although more recent work suggests that when clear scoring criteria are provided, better levels of agreement between assessors are achieved (King & Green 2011a; Phythian *et al* 2012).

Clinical observations, such as the assessment of lameness in sheep, are diagnostic tests. Accordingly, the principles used to validate diagnostic tests are used to assess the test validity. For diagnostic test evaluation, validity is defined as the ability of a test to produce correct test results compared to a reference or 'gold standard' (Greiner & Gardner 2000). Since there is no current gold standard for lameness scoring in sheep, validity can only be assessed indirectly. Convergent validity concerns the correlation between tests for the same condition, for example the correlation between different methods of lameness assessment.

Reliability is also an important aspect of test validity (Abramson & Abramson 2008). Therefore, the between- and within-observer reliability of both lameness scoring systems for sheep under examination here have previously been determined (Phythian 2011; Phythian *et al* 2012).

As well as being valid and reliable, a diagnostic test must be feasible for use and applicable under different conditions (Knierim & Winckler 2009). Since the management and environmental conditions for assessment may vary between flocks it may not be feasible to apply a complex multiple category lameness scoring system under all farm systems or terrains. Therefore, a simplified binary scoring system that allows the assessor to walk amongst groups of sheep and record whether

Figure 1



Prevalence of lameness recorded on each study farm (n = 40) by the group observation method (GOM) and individual animal gait assessment (IAGA). Farms in which GOM identified lame sheep but IAGA identified zero lame sheep are denoted by the symbol (*). The reference line denotes the threshold of 5% lameness prevalence.

individual animals are 'lame' or 'sound', could provide a more practical measurement for on-farm welfare assessments.

The objective of this study was to firstly investigate the convergent validity, or agreement between, two on-farm lameness assessment methods; a binary lameness scoring system applied to animals at pasture (group observation method; GOM) and an individual animal gait assessment (IAGA) in penned sheep. Secondly, the association between the level of lameness observed by IAGA and the presence of specific foot lesions recorded on clinical examination was evaluated.

Materials and methods

Study population

The investigation was a cross-sectional study conducted during July 2008–May 2009 on 40 farms in Northern England and Wales. Farms were recruited through contact with their local veterinary practices. The inclusion criteria were the informed and written consent to participate, and the distance of the farm from the University of Liverpool, School of Veterinary Science, Leahurst, Wirral (< 150-mile radius). Study farms were classified as lowland (n = 18), upland (n = 11) and hill (n = 11) farm types. Since the objective of the study was to validate testing protocols and not to assess farm-level prevalence of lameness, farmers were requested to present a sample of approximately 100 sheep including as many lame sheep as could be conveniently gathered.

Methods of lameness assessment

The group observation method (GOM) involved walking the group of sheep around the field or paddock for 10 to 25 min, depending on the number of sheep presented, to facilitate the identification and counting of the number of lame animals in the group (Phythian *et al* 2012). In this method, the identity of individual lame sheep could not be recorded. Lameness was defined as the observation of one or more of the following signs: visible nodding of head in time with a short stride; grazing 'on knees'; uneven gait; arching of the back during locomotion; non-weight bearing on an affected limb; extreme difficulty rising; reluctance to move once standing, ie a lame sheep in the present study was equivalent to a gait score ≥ 2 as per Kaler *et al* (2009).

For the individual animal gait assessment (IAGA), the sample group was gathered to a holding area and the gait of each individual sheep was observed by walking the sheep around an assessment pen (approximately 2 × 2 m) with a level flooring surface, relatively free of dirt, debris and bedding. Individual sheep were then walked around the pen for 1 to 2 min to examine the gait in both directions and scored as either 'sound' or 'lame' according to the criteria listed above. The observer then examined the feet of each sheep for the presence of white line lesions (WL) — separation and detachment of the white line ('shelly hoof'), including impaction and infection, inter-digital dermatitis

https://doi.org/10.7120/09627286.22.4.417 Published online by Cambridge University Press

(IDD), footrot (FR), contagious ovine digital dermatitis (CODD), toe granuloma (TG), and joint swellings (JS) as described by Winter (2001). All assessments were performed by the same veterinary assessor.

The study was approved by the University of Liverpool Ethics Committee (RETH000287).

Data analysis

Data were analysed using Stata version 10 (StataCorp LP, College Station, TX, USA). The level of agreement between the two tests was investigated using simple linear regression of the percentage of sheep identified as lame by GOM and IAGA. The prevalence of each lesion by the two levels of the individual animal gait assessment method (sound and lame) was examined. The level of test agreement was also examined using Bland Altman plots (Bland & Altman 1986) in which the difference between the tests was plotted against the mean value, and shown within 2 standard deviations (SD) known as the 'lines of equality'.

The percentage of lameness in each sample group was also stratified into 'high prevalence' (> 5%) and 'low prevalence' (< 5%) according to FAWC recommendations (FAWC 2011), allowing the tests to be evaluated in terms of their ability to identify high and low prevalence flocks.

Results

The mean percentage of lameness in the presented sample of 3,094 sheep from 40 farms recorded by group observation (GOM) was 7.1% (5.5–8.8%), whilst the mean percentage found by individual examination (IAGA) was 6.2% (4.3–8.1%). No significant difference between the percentage of lameness identified by GOM and IAGA was found. The percentage of sheep identified as lame on each farm is represented in Figure 1.

Almost half of the sample population was observed with a foot lesion present in one or more feet (44%) — the most frequently recorded lesion was white line disease (WL) 43.9%. Table 1 shows the remaining 1% of sheep were identified with inter-digital dermatitis (IDD), 1.3% with footrot (FR), 1.0% contagious ovine digital dermatitis (CODD), 1.1% toe granuloma (TG), and 0.5% with joint swellings (JS). Eighty-five percent of sheep with IDD and 80% of sheep with FR were identified as lame by the group observation method. However, only 11.6% of sheep with WL, 61.5% with IDD, 59% with FR and 42% with TG were found to be lame. By contrast, 83.9% with CODD and 71.4% with JS were found to be lame during individual gait assessment (Table 1). A further 3.6% of animals identified lame on IAGA had no clinically detectable lesions.

Visual examination of Figure 1 suggests that 75% of sample groups with a high prevalence of lameness (> 5%) would be detected using GOM and IAGA. Overall, the level of agreement between the two tests was good (R^2 = 55.8%; Figure 2), and stratification of the data (Table 2) identified that GOM and IAGA would place presented samples into the same categories (high or low lameness prevalence) in 87.5% of assessments. Bland-Altman analysis demonstrated that the percentage of lame sheep observed by the two methods mostly fell within the lines of equality shown as the two reference lines on Figure 3, indicating good agreement between the two methods. However, there were eight assessments in which the sample group was identified with low lameness prevalence (≤ 5%) by GOM, but no signs of lameness were detected by IAGA (Figure 1).

Table 1 The number and percentage of sheep with foot lesions detected as lame by individual animal gait assessment (IAGA).

Foot lesion	Number (%) observed with foot lesion	Number (%) with lesion observed as lame
WL	1,350 (43.9)	157 (11.6)
IDD	26 (0.8)	16 (61.5)
FR	39 (1.3)	23 (59.0)
CODD	31 (1.0)	26 (83.9)
TG	34 (1.1)	14 (41.2)
JS	14 (0.5)	10 (71.4)

WL: white line lesion; IDD: inter-digital dermatitis; FR: footrot; CODD: contagious ovine digital dermatitis; TG: toe granuloma; JS: joint swellings.

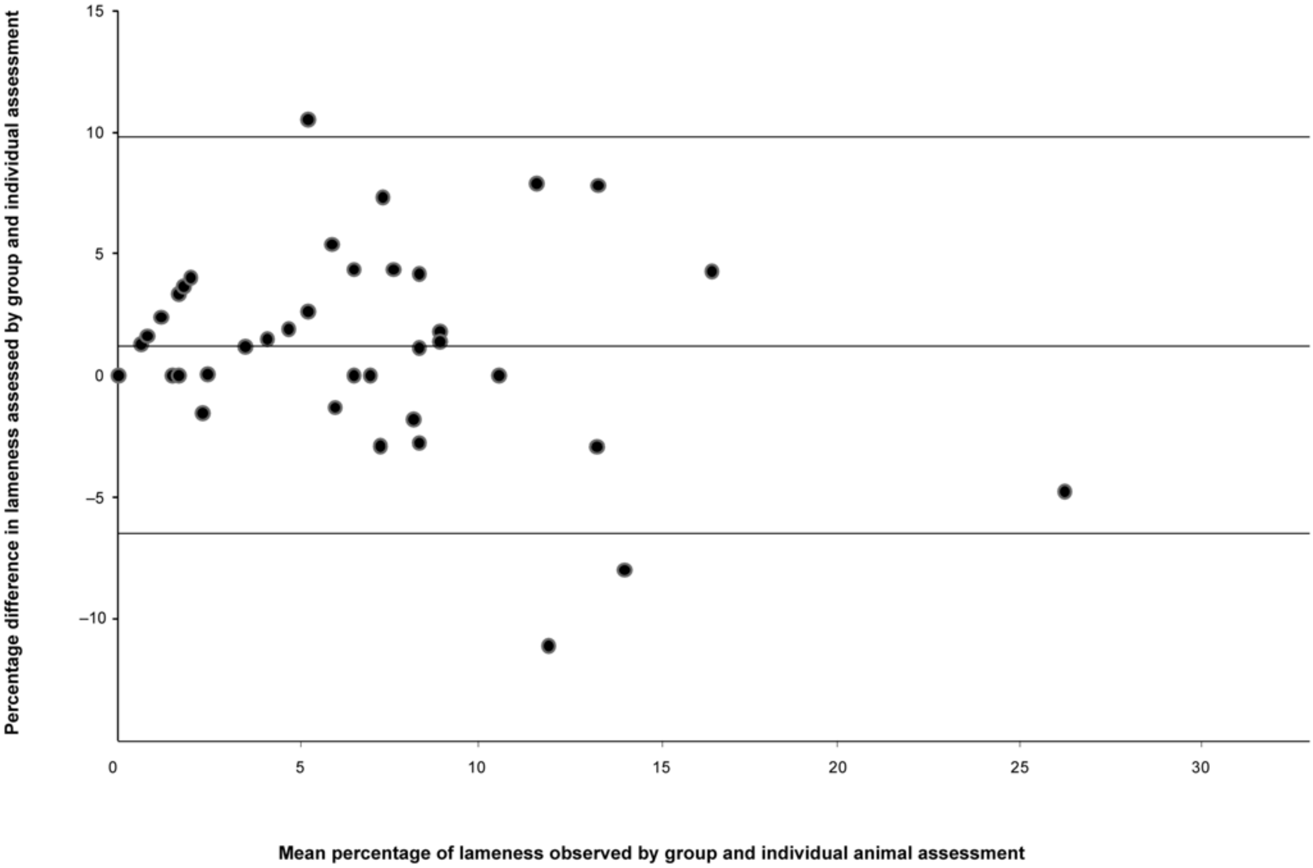
Discussion

This study demonstrates a high level of agreement or convergent validity between the two methods of lameness assessment. Together with the inter-observer reliability estimates for these lameness assessment methods in sheep (Phythian *et al* 2012), the results of this study provide confidence in their validity as diagnostic tests for assessing lameness in sheep.

Comparison of lameness prevalence identified by two assessment methods

The prevalence and severity of lameness is regularly measured as part of routine stockperson assessments, veterinary examinations, farm assurance and statutory welfare inspections of sheep flocks. The objective of many animal welfare assessments is to improve or maintain good standards of animal welfare. Benchmarks and target standards are often used as part of animal welfare monitoring systems in order to provide the producer with feedback and advice where necessary. There are no widely accepted benchmarks for sheep lameness, but recently FAWC (2011) suggested that the UK flock should aim for a prevalence of 5% or less by 2016. Therefore, a level of 5% was set as the threshold between 'high' and 'low' lameness categories and used to compare the performance of the two assessment methods. This approach demonstrated that, for the majority of study samples, the two lameness methods produced the same result and assigned the presented groups of sheep into the same lameness category.

Figure 2



Bland-Altman plot of the difference between the mean proportion of lame sheep by the group observation method (GOM) and individual animal gait assessment (IAGA). Solid circles represent a farm assessment. The reference lines denote 2 standard deviations from the mean.

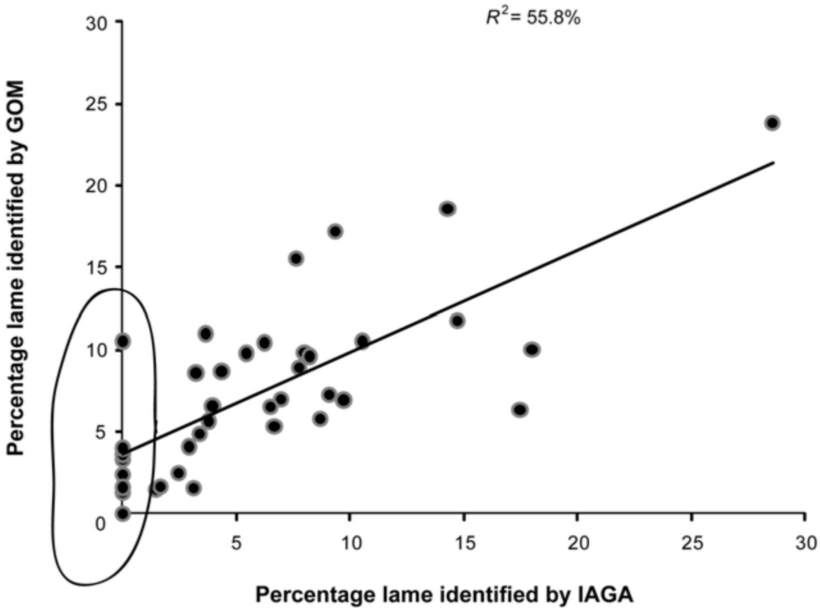
Table 2 A comparison of the prevalence of lameness (as assessed by group and individual assessment methods) when a threshold of 5% is used to classify all farms into high or low lameness categories.

		Group assessment method	
		High lameness category (> 5%)	**Low lameness category (≤ 5%)**
Individual assessment	**High lameness category (> 5%)**	n = 19, 47.5%	n = 0, 0%
	Low lameness category (≤ 5%)	n = 5, 12.5%	n = 16, 40.0%

However, on a number of farms (n = 8), sheep were identified as lame on group assessment but were not identified as lame when the same animals were individually gait scored. This may be a result of the stress of gathering of the sheep and their close proximity to human observers post gathering which could have altered their behavioural expressions (Fitzpatrick *et al* 2006). Consequently, changes in behaviour may have meant that animals clearly observed to be lame on group observation may not have been observed as lame on individual animal examination. Additionally, as expected, the quality of the examination area available was variable between farms. Individual gait assessment was difficult to perform in poorly lit, straw-bedded and circular assessment areas present on some farms and was facilitated by gathering the sheep into a holding pen, then conducting examinations in a well-lit, rectangular-shaped pen with a non-slip and clean floor, free from debris or bedding. Group observation has been demonstrated to be a reliable method of assessing lameness (Phythian *et al* 2012). On-farm findings suggest that group observation is a more feasible and quicker method than individual animal assessment for assessing farms with ≤ 5% lameness, and this study further supports use of group observation as a valid and reliable lameness screening tool for on-farm welfare inspections.

https://doi.org/10.7120/09627286.22.4.417 Published online by Cambridge University Press

Figure 3



Association between the percentage of lame sheep on each farm diagnosed by group observation method (GOM) and individual animal gait assessment (IAGA). Solid black circles indicate a farm assessment. Farms in which sheep were identified to be lame by GOM but were not identified lame on IAGA are shown within the open oval.
The reference line denotes the best line of fit.

Study population

Farmers were asked to select groups with a high prevalence of lameness. However, the prevalence of lameness identified in this sample by the two assessment methods is only slightly lower (7.1 and 6.2%) than the 10.8% prevalence of lameness reported by farmers in England in a recent survey (Kaler & Green 2008a), and is very close to the 6.9% prevalence of sheep with gait scores ≥ 2 identified during a recent research study (King & Green 2011b), suggesting that the study sample is reasonably representative of English flocks in terms of lameness prevalence and was a suitable sample to use to test the lameness-scoring methods.

Relationship between lameness scores and foot lesions

In agreement with the study by Conington *et al* (2010), nearly half of the animals in this study were identified with WL, although this lesion did not appear to be associated with lameness on most farms. Indeed, the level of lameness in sheep observed with WL was relatively low at 11.6%. Therefore, an assessment of lameness using the two methods described here could not be expected to indicate whether sheep are affected by white line disease. This would require examination of feet for the presence of foot lesions. Although not formally quantified in this study, subjective experience from this study (and in agreement with Winter 2008) suggests that separation of the white line alone was not associated with lameness, but impaction and infection of the separated area could result in lameness. Clearly, this is an area that requires further investigation, for example an examination of gross and histopathological aspects of foot lesions along with the application of categorical lameness and pain-scoring methods to explore the relationship between the intensity of the pain response and the pathology of white line lesions.

In contrast to the findings with WL, there was a good association between individual lameness assessment and the presence of CODD and JS with 83.9 and 71.4% of sheep, respectively, with these foot lesions found to be lame. This is likely to reflect the severity of the foot and joint pathology and the degree of pain associated with these conditions resulting in sheep exhibiting severe signs of lameness during individual animal assessment. Surprisingly, there was not as strong an association between lameness and FR and IDD scores, which may be a result of the very low prevalence of footrot in the sample population (1.0%) and the fact that unlike other studies (Ley *et al* 1989) no staging of footrot lesions was performed. Therefore, whilst footrot is well recognised as a painful disease for sheep (Ley *et al* 1989; Fitzpatrick *et al* 2006), it might be expected that some sheep in the present study were diagnosed with healing footrot lesions. In these instances, it is possible that the underlying inflammation and pathology was resolving thus resulting in very mild signs of lameness that were not detectable during individual animal gait assessment. Again, this is an area that warrants further investigation.

Feasibility

Following pilot study testing of categorical lameness scoring scales (CJ Phythian, unpublished observations 2008), a simple binary scoring system, based on the descriptors of Kaler *et al* (2009) was used here since this approach to gait assessment was found to be the most reliable and feasible method of assessment. It was recognised that a disadvantage of using a binary scoring scale in this study was that the association between the severity of lameness and foot lesions was not examined. Recent work by King and Green (2011a), using a categorical lameness scoring system, has identified an association between the lameness scores and the severity of foot lesions in sheep. Indeed, such a system would be advantageous for detailed monitoring of, for example, treatment outcomes or for research purposes. However, for welfare assessment and estimation of lameness prevalence by farmers and veterinary surgeons, a simple binary system, in which sheep were scored as either 'sound' or 'lame' regardless of the severity of the gait, may be a more feasible, but still adequate way, of identifying lame sheep and thus enable the employment of treatment and control measures.

The method of group observation generally works well for assessing animals maintained in groups of 24–120 at pasture (Phythian *et al* 2012) irrespective of terrain. Whilst the method was found to be feasible in this study, it is not known whether larger group sizes (n > 120) can be reliably assessed in this way. It is suggested that assessments of sheep widely dispersed or located in rugged and steep terrains or those that move some distance away from the observer may be facilitated by the use of shepherding dogs, farmer involvement or use of an all-terrain vehicle.

Animal welfare implications

This study suggested that a binary scoring scale for assessing lameness was valid and feasible to apply on commercial farms in which sheep were managed under different farming systems. These results could inform the methods for the monitoring and surveillance of sheep lameness as part of veterinary flock health and welfare planning and lameness prevention and control programmes, and for on-farm welfare assessments performed by farm animal welfare inspectors.

Conclusion

The high level of agreement found between the percentage of lameness observed by the two assessment methods provides additional evidence for their validity as diagnostic tests of lameness in sheep. The group observation method appears to be a more feasible approach than individual animal gait assessment when applied under field conditions. Furthermore, applying lameness cut-off values of 5% to farm assessments showed that the majority of study farms would have been assigned the same benchmark or lameness prevalence category by both gait assessment methods.

Acknowledgements

This work was funded Defra as part of the AW1025 project (Development of indicators for the on-farm assessment of sheep welfare). None of the authors has any financial or personal relationships that could inappropriately influence or bias the content of the paper.

References

Abramson ZH and Abramson JH 2008 *Research Methods in Community Medicine: Surveys, Epidemiological Research, Programme Evaluation, Clinical Trials* pp 161-178. John Wiley and Sons: Chichester, UK

Bland JM and Altman DG 1986 Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet 1*: 307-310. http://dx.doi.org/10.1016/S0140-6736(86)90837-8

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

Farm Animal Welfare Council (FAWC) 2011 *Opinion on Lameness in Sheep*. FAWC: London, UK. http://www.fawc.org.uk /pdf/sheep-lameness-opinion-110328.pdf

Fitzpatrick J, Scott M and Nolan A 2006 Assessment of pain and welfare in sheep. *Small Ruminant Research 62*: 55-61. http://dx.doi.org/10.1016/j.smallrumres.2005.07.028

Greiner M and Gardner A 2000 Application of diagnostic tests in veterinary epidemiologic studies. *Preventive Veterinary Medicine 45*: 43-59. http://dx.doi.org/10.1016/S0167-5877(00)00116-1

Harkins LS 2005 *Development of a prototype welfare tool for use in sheep*. Master of Science Thesis, University of Glasgow, Glasgow, UK

Hodgkinson O 2010 The importance of feet examination in sheep health management. *Small Ruminant Research 92*: 67-71. http://dx.doi.org/10.1016/j.smallrumres.2010.04.007

Kaler J and Green LE 2008a Naming and recognition of six foot lesions of sheep using written and pictorial information: a study of 809 English sheep farmers. *Preventive Veterinary Medicine 83*: 52-64. http://dx.doi.org/10.1016/j.prevetmed.2007.06.003

Kaler J and Green LE 2008b Recognition of lameness and decisions to catch for inspection among sheep farmers and specialists in GB. *BMC Veterinary Research 4*: 41. http://dx.doi.org/10.1186/1 746-6148-4-41

Kaler J, Wassink GJ and Green LE 2009 The inter- and intra-observer reliability of a locomotion scoring scale for sheep. *The Veterinary Journal 180*: 189-194. http://dx.doi.org/10.1016/ j.tvjl.2007.12.028

King EM and Green LE 2011a Why are sheep lame? Temporal associations between severity of foot lesions and severity of lameness in 60 sheep. *Animal Welfare 20*: 433-438

King EM and Green LE 2011b Assessment of farmer recognition and reporting of lameness in adults in 35 lowland sheep flocks in England. *Animal Welfare 20*: 321-328

Knierim U and Winckler C 2009 On-farm welfare assessment in cattle: validity, reliability and feasibility issues and future perspectives with special regard to the Welfare Quality® approach. *Animal Welfare 18*: 451-458

Ley SJ, Livingston A and Waterman AE 1989 The effect of chronic clinical pain on thermal and mechanical thresholds in sheep. *Pain 39*: 353-357. http://dx.doi.org/10.1016/0304-3959(89)90049-3

Nieuwhof GJ and Bishop SC 2005 Costs of the major endemic diseases of sheep in Great Britain and the potential benefits of reduction in disease impact. *Animal Science 81*: 23-29. http://dx.doi.org/10.1079/ASC41010023

Phythian CJ 2011 *Development of indicators for the on-farm assessment of sheep welfare*. PhD Thesis, University of Liverpool, Liverpool, UK

Phythian CJ, Cripps PJ, Michalopoulou E, Jones PH, Grove-White D, Clarkson MJ, Winter AC, Stubbings LA and Duncan JS 2012 Reliability of on-farm indicators of sheep welfare assessed by a group observation method. *The Veterinary Journal 193*: 257-263. http://dx.doi.org/10.1016/j.tvjl.2011.12.006

Winter AC 2001 Lameness in sheep: 1. Diagnosis. *In Practice 26*: 58-63. http://dx.doi.org/10.1136/inpract.26.2.58

Winter AC 2008 Lameness in sheep. *Small Ruminant Research 76*: 149-153. http://dx.doi.org/10.1016/j.smallrumres.2007.12.008

https://doi.org/10.7120/09627286.22.4.417 Published online by Cambridge University Press