A practical tool for locomotion scoring in sheep: Reliability when used by veterinary surgeons and sheep farmers

Article *in* The Veterinary record · February 2015

DOI: 10.1136/vr.102882 · Source: PubMed

CITATIONS
12

READS
180

4 authors:

Joseph William Angell
Wern Veterinary Surgeons
58 PUBLICATIONS **141** CITATIONS
SEE PROFILE

Peter Cripps
University of Liverpool
234 PUBLICATIONS **4,776** CITATIONS
SEE PROFILE

Dai H Grove-White
University of Liverpool
89 PUBLICATIONS **958** CITATIONS
SEE PROFILE

Jennifer Sarah Duncan
University of Liverpool
80 PUBLICATIONS **484** CITATIONS
SEE PROFILE

Original Article

A practical tool for locomotion scoring in sheep: its reliability when used by veterinary surgeons and sheep farmers

J.W. Angell BVSc MRCVS[a,*],
P.J. Cripps BSc(Hons) BVSc MSc PhD MRCVS [a],
D.H. Grove-White BVSc MSc DBR PHD DipECPHM FRCVS [a],
J.S. Duncan BSc(Hons) BVSc DipECSRHM PhD MRCVS[a]

[a]Veterinary Epidemiology, Institute of Infection and Global Health, The University of Liverpool, Leahurst Campus, Neston, Wirral, CH64 7TE

*Corresponding author. Tel.: +44 151 7946050; Fax: +44 151 7946034; E-mail address: jwa@liv.ac.uk

Abstract

A four-point locomotion scoring tool for sheep was developed and tested on 10 general practice veterinary surgeons (VS) and 10 sheep farmers. Thirty-four video clips of sheep displaying different locomotion scores were recorded and randomly assorted. Following a set period of training using four other video clips typical of the four locomotion scores, participants then scored the 34 test clips. The participants repeated the training and the exercise one month later. There were high levels of intra-observer repeatability: weighted kappa (κ_W) 0.81 for VS and 0.83 for farmers. There was no difference in intra-observer repeatability between vets and farmers (Wilcoxon signed rank P = 0.8). When considering the overall distribution of scores within the video-package, there were high levels of inter-observer repeatability: mean κ_W 0.73 for VS and 0.72 for farmers. However, the repeatability for the individual locomotion scores was only fair to moderate. It is therefore recommended that when observations are repeated on different occasions they are made by the same observer.

Keywords: Lameness; Locomotion; Mobility; Scoring; Repeatability; Reliability; Sheep

Introduction

Lameness in sheep is a priority welfare concern for the UK sheep industry (Phythian and others 2011). Current understanding and consensus of opinion have led to recommendations to farmers to identify and treat lame sheep early (FAWC 2011).

Locomotion scoring identifies lame individuals and can be used to determine flock prevalence. Previous scoring tools e.g. Kaler and others (2009); Ley and others (1989); Phythian and others (2012) and Welsh and others (1993) all have limitations, being either overly detailed or simplistic. Furthermore, they all use small numbers of experienced researchers, which could reduce generalisability.

The aim of this study was to develop a locomotion scoring tool for use by farmers, and veterinary surgeons (VS) to assess lameness severity in individual sheep, and severity and prevalence in flocks.

Materials and Methods

Locomotion Scoring

A four-point system was developed by combining the Kaler system (Kaler and others 2009) with the DairyCo Mobility Scoring System (DairyCo 2009):

0: (SOUND) Bears weight evenly on all four feet and walks with an even rhythm.

1: (MILDLY LAME) Steps are uneven but it is not clear which limb or limbs are affected.

2: (MODERATELY LAME) Steps are uneven and the stride may be shortened; the affected limb or limbs are identifiable.

3: (SEVERELY LAME) Mobility is severely compromised such that the sheep frequently stops walking or lies down due to obvious discomfort. The affected limb or limbs are clearly identifiable and may be held off the ground whilst walking or standing.

Ethical approval was provided by the University of Liverpool (VREC13).

Thirty-eight video clips of sheep walking and standing were made – representing all four scores. To ensure a range of severities was represented, these were scored by three experienced sheep VS to collectively determine the 'true' score. Four of the clips were used to train the participants. The other 34 were shown in random order to the participants. If there was more than one sheep visible, a red circle was drawn around the relevant individual.

Study Population

The tool was tested on a convenience, non-random sample of 10 general practice VS and 10 sheep farmers.

Training

Each participant was trained using clips typical of each score. Participants then watched the test clips taking as long as needed and were allowed to watch each clip as many times as necessary. No help was given during the test period. The training and assessment were repeated one month later.

Data Analysis

Intra-observer agreement

a) Bias between attempts

For each clip, the score from an individual's second attempt was subtracted from their initial score. Differences were investigated using one-sample t-tests.

b) Exact agreement

Percent agreement for an observer was determined from the number of observations that matched exactly:

$$\frac{(\text{number of matching observations})}{34} \times 100$$

The mean percent agreement was calculated for VS and farmers, and compared using the Chi-squared test. Similar data were compiled for the one and two point differences.

c) Pairwise Kappa

Weighted Kappa (κ_W) was calculated between each pair of observations by each observer using quadratic weights and interpreted using Landis and Koch (1977), Table 2.

Inter-observer agreement

a) Kappa between observers

For each observer, a κ_W was created with each member of their group. The mean of these nine values was that individual's inter-observer agreement.

b) Kappa for locomotion scores

To examine the repeatability of recording different severities of locomotion score, unweighted κ was obtained for all clips that had received the given score.

c) Median scores

Median scores for each clip were calculated for both VS and farmers. Differences were assessed using the Wilcoxon signed-rank test.

Statistical significance was set at <0.05. All analyses used STATA13 (StataCorp, Texas).

Results

All participants found the tool easy to use. They found it hardest to distinguish between scores 1 and 0.

The mean proportion of scores attributed from the first set of observations was: score 0: 8.7 (26%), score 1: 9.9 (29%), score 2: 9.8 (29%) and score 3: 5.7 (17%).

Intra-observer agreement (Table 1)

Individual Observers	Locomotion Score Mean (SD)	Mean observer difference in locomotion scores between observations	t-test of mean observer difference compared to zero (P value)	Intra-observer agreement (%)			Intra-observer κ_W for each individual observer comparing first and second observations	Inter-observer agreement: Mean κ_w for each observer versus all 9 other observers in group
				Exact agreement	+/- 1 point difference	+/- 2 point difference		
Veterinary Surgeon Observers								
Vet 1	1.31 (1.16)	-0.12	0.013	61.8	100	100	0.86	0.78
Vet 2	1.37 (1.09)	-0.06	0.344	79.4	100	100	0.91	0.75
Vet 3	1.56 (1.15)	0.13	0.076	70.6	100	100	0.89	0.75
Vet 4	1.46 (0.97)	0.03	0.647	55.9	94.1	100	0.67	0.72
Vet 5	1.22 (1.13)	-0.21	0.003	79.4	100	100	0.92	0.76
Vet 6	1.38 (1.07)	0.19	0.001	61.8	97.1	100	0.74	0.74
Vet 7	1.41 (1.03)	-0.02	0.762	61.8	97.1	100	0.77	0.76
Vet 8	1.41 (1.03)	-0.02	0.825	55.9	97.1	100	0.75	0.64
Vet 9	1.54 (0.97)	0.11	0.129	58.8	97.1	100	0.73	0.68
Vet 10	1.41 (1.03)	-0.02	0.782	64.7	97.1	100	0.83	0.70
Overall Mean (SD)		**-0.01 (0.05)**		**65.0 (8.7)**	**98.0 (2.0)**		**0.81 (0.1)**	**0.73 (0.04)**
Farmer Observers								
Farmer 1	1.04 (0.95)	-0.27	<0.001	79.4	100	100	0.89	0.71
Farmer 2	1.26 (1.07)	-0.05	0.530	52.9	94.1	100	0.72	0.67
Farmer 3	1.63 (1.16)	0.32	<0.001	64.7	97.1	100	0.83	0.70
Farmer 4	1.31 (1.22)	0.00	0.952	70.6	97.1	100	0.87	0.77
Farmer 5	1.56 (0.92)	0.25	<0.001	76.5	100	100	0.86	0.69
Farmer 6	1.43 (1.01)	0.11	0.026	61.8	100	100	0.81	0.73
Farmer 7	1.26 (1.05)	-0.05	0.306	67.7	97.1	100	0.81	0.77
Farmer 8	1.07 (1.11)	-0.24	<0.001	91.2	100	100	0.96	0.78
Farmer 9	1.22 (1.01)	-0.09	0.198	67.7	94.1	100	0.75	0.70
Farmer 10	1.34 (1.05)	0.35	0.678	50.0	100	100	0.77	0.73
Overall Mean (SD)		**0.00 (0.06)**		**68.3 (12.2)**	**98.0 (2.4)**		**0.83 (0.1)**	**0.72 (0.04)**
Comparison of difference in means between VS and Farmer observers (P value)				**0.5**	**1.0**		**0.8**	**0.8**

Table 1: Intra- and Inter-observer agreement for veterinary surgeon and farmer observers.

184 *a) Bias between attempts*

185 Bias was present within and between observers and was significant for three
186 VS and five farmers. The largest differences in scores were -0.21 and 0.25
187 respectively.

188

189 *b) Exact agreement*

190 The mean overall exact agreement within individual observers was 65.0%
191 (SD 8.7) for VS and 68.3% (SD 12.2) for farmers (P = 0.5).

192

193 *c) Pairwise kappa*

194 The mean κ_W at intra-observer level was 0.81 for VS and 0.83 for farmers (P
195 = 0.8).

196

197 *Inter-observer agreement*

198 *a) Kappa between observers (Table 1)*

199 The mean κ_w at inter-observer level was 0.73 (SD 0.04) for VS and 0.72 (SD
200 0.04) for farmers (P = 0.8).

201

202 *b) Kappa for locomotion scores (Table 2)*

203 Overall, for score 3 there is substantial agreement between observers. For
204 other scores, there is moderate or fair agreement.

205

206 *c) Median scores*

207 The median score assigned to each video clip by VS was not significantly
208 different to that assigned by farmers (P = 0.18) (Table 2).

Locomotion score	Overall κ VS and Farmers[§]	Overall κ VS[§]	Overall κ Farmers[§]
0	0.37 Fair	0.30 Fair	0.43 Moderate
1	0.22 Fair	0.27 Fair	0.18 Slight
2	0.43 Moderate	0.47 Moderate	0.41 Moderate
3	0.62 Substantial	0.67 Substantial	0.58 Moderate
Combined	**0.40 Fair**	**0.42 Moderate**	**0.39 Fair**

Table 2: Inter-observer agreement for individual locomotion scores

[§] Interpretations are taken from Landis and Koch, (1977): 0 = poor; 0.01 to 0.20 = slight; 0.21 to 0.40 = fair; 0.41 to 0.60 = moderate; 0.61 to 0.80 = substantial; 0.81 to 1.00 = almost perfect.

Discussion

There were score differences between observation attempts, however we consider this bias, whilst present, is too small to invalidate the scoring system. The variation in locomotion scores (Table 2) indicates bias between observers and may have led to smaller κ values than if the scores had equal prevalence within the video package (Byrt and others 1993). However, given that the lowest prevalence score (score 3) had the highest levels of repeatability between observers, a more equal prevalence would likely have had little impact on the κ values. Both intra- and inter-observer repeatability were

substantial indicating that this tool could be used reliably in monitoring lameness in individuals over time and enable different observers to reliably measure lameness across farms. However, the inter-observer repeatability of locomotion scores was slight to moderate, except for score 3. Therefore, whilst different observers scored similar proportions of sheep with each locomotion score, the ability to score the same individual with the same score was unsatisfactory.

The large number and two types of observers in this study suggest that the tool is applicable to industry users.

Conflicts of Interest

None of the authors of this paper has a financial or personal relationship with other people or organisations that could inappropriately influence or bias the content of the paper.

Acknowledgments

This study was supported by a grant from the British Veterinary Association Animal Welfare Foundation, from the Norman Hayward Fund, and also by a grant from Hybu Cig Cymru/Meat Promotion Wales. The authors are grateful to all the veterinary surgeons and farmers who willingly agreed to take part, including those who willingly provided their sheep.

References

BYRT, T., BISHOP, J. & CARLIN, J. B. (1993) Bias, prevalence and kappa. J Clin Epidemiol 46, 423-429

DAIRYCO (2009) Mobility Scoring. In Mobility scoring tool, DairyCo

FAWC (2011) Farm Animal Welfare Council - opinion on lameness in sheep. In Farm Animal Welfare Council Reports

KALER, J., WASSINK, G. J. & GREEN, L. E. (2009) The inter- and intra-observer reliability of a locomotion scoring scale for sheep. Vet J 180, 189-194

LANDIS, J. R. & KOCH, G. G. (1977) The measurement of observer agreement for categorical data. Biometrics 33, 159-174

LEY, S. J., LIVINGSTON, A. & WATERMAN, A. E. (1989) The effect of chronic clinical pain on thermal and mechanical thresholds in sheep. Pain 39, 353-357

PHYTHIAN, C. J., CRIPPS, P. J., MICHALOPOULOU, E., JONES, P. H., GROVE-WHITE, D., CLARKSON, M. J., WINTER, A. C., STUBBINGS, L. A. & DUNCAN, J. S. (2012) Reliability of indicators of sheep welfare assessed by a group observation method. The Veterinary Journal 193, 257-263

PHYTHIAN, C. J., MICHALOPOULOU, E., JONES, P. H., WINTER, A. C., CLARKSON, M. J., STUBBINGS, L. A., GROVE-WHITE, D., CRIPPS, P. J. & DUNCAN, J. S. (2011) Validating indicators of sheep welfare through a consensus of expert opinion. Animal 5, 943-952

WELSH, E. M., GETTINBY, G. & NOLAN, A. M. (1993) Comparison of a visual analogue scale and a numerical rating scale for assessment of lameness, using sheep as a model. Am J Vet Res 54, 976-983