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

Peter Cripps
University of Liverpool
234 PUBLICATIONS 4,776 CITATIONS

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

Jennifer Sarah Duncan
University of Liverpool
80 PUBLICATIONS 484 CITATIONS

Some of the authors of this publication are also working on these related projects:

On-farm assessment of sheep welfare View project
Research projects ideas development. View project
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\textsuperscript{a,\#},
P.J. Cripps BSc(Hons) BVSc MSc PhD MRCVS\textsuperscript{a},
D.H. Grove-White BVSc MSc DBR PHD DipECPHM FRCVS\textsuperscript{a},
J.S. Duncan BSc(Hons) BVSc DipECSRHM PhD MRCVS\textsuperscript{a}

\textsuperscript{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 ($\kappa_W$) 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 $\kappa_W$ 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:

\[
\text{(number of matching observations)} \times 100
\]

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

<table>
<thead>
<tr>
<th>Individual Observers</th>
<th>Locomotion Score Mean (SD)</th>
<th>Mean observer difference in locomotion scores between observations</th>
<th>t-test of mean observer difference compared to zero (P value)</th>
<th>Intra-observer agreement (%)</th>
<th>Intra-observer ( \kappa_w ) for each individual observer comparing first and second observations</th>
<th>Inter-observer agreement: Mean ( \kappa_w ) for each observer versus all 9 other observers in group</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Veterinary Surgeon Observers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vet 1</td>
<td>1.31 (1.16)</td>
<td>-0.12</td>
<td>0.013</td>
<td>61.8</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Vet 2</td>
<td>1.37 (1.09)</td>
<td>-0.06</td>
<td>0.344</td>
<td>79.4</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Vet 3</td>
<td>1.56 (1.15)</td>
<td>0.13</td>
<td>0.076</td>
<td>70.6</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Vet 4</td>
<td>1.46 (0.97)</td>
<td>0.03</td>
<td>0.647</td>
<td>55.9</td>
<td>94.1</td>
<td>100</td>
</tr>
<tr>
<td>Vet 5</td>
<td>1.22 (1.13)</td>
<td>-0.21</td>
<td>0.003</td>
<td>79.4</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Vet 6</td>
<td>1.38 (1.07)</td>
<td>0.19</td>
<td>0.001</td>
<td>61.8</td>
<td>97.1</td>
<td>100</td>
</tr>
<tr>
<td>Vet 7</td>
<td>1.41 (1.03)</td>
<td>-0.02</td>
<td>0.762</td>
<td>61.8</td>
<td>97.1</td>
<td>100</td>
</tr>
<tr>
<td>Vet 8</td>
<td>1.41 (1.03)</td>
<td>-0.02</td>
<td>0.825</td>
<td>55.9</td>
<td>97.1</td>
<td>100</td>
</tr>
<tr>
<td>Vet 9</td>
<td>1.54 (0.97)</td>
<td>0.11</td>
<td>0.129</td>
<td>58.8</td>
<td>97.1</td>
<td>100</td>
</tr>
<tr>
<td>Vet 10</td>
<td>1.41 (1.03)</td>
<td>-0.02</td>
<td>0.782</td>
<td>64.7</td>
<td>97.1</td>
<td>100</td>
</tr>
<tr>
<td><strong>Overall Mean (SD)</strong></td>
<td>-0.01 (0.05)</td>
<td></td>
<td></td>
<td>65.0 (8.7)</td>
<td>98.0 (2.0)</td>
<td></td>
</tr>
<tr>
<td><strong>Farmer Observers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Farmer 1</td>
<td>1.04 (0.95)</td>
<td>-0.27</td>
<td>&lt;0.001</td>
<td>79.4</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Farmer 2</td>
<td>1.26 (1.07)</td>
<td>-0.05</td>
<td>0.530</td>
<td>52.9</td>
<td>94.1</td>
<td>100</td>
</tr>
<tr>
<td>Farmer 3</td>
<td>1.63 (1.16)</td>
<td>0.32</td>
<td>&lt;0.001</td>
<td>64.7</td>
<td>97.1</td>
<td>100</td>
</tr>
<tr>
<td>Farmer 4</td>
<td>1.31 (1.22)</td>
<td>0.00</td>
<td>0.952</td>
<td>70.6</td>
<td>97.1</td>
<td>100</td>
</tr>
<tr>
<td>Farmer 5</td>
<td>1.56 (0.92)</td>
<td>0.25</td>
<td>&lt;0.001</td>
<td>76.5</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Farmer 6</td>
<td>1.43 (1.01)</td>
<td>0.11</td>
<td>0.026</td>
<td>61.8</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Farmer 7</td>
<td>1.26 (1.05)</td>
<td>-0.05</td>
<td>0.306</td>
<td>67.7</td>
<td>97.1</td>
<td>100</td>
</tr>
<tr>
<td>Farmer 8</td>
<td>1.07 (1.11)</td>
<td>-0.24</td>
<td>&lt;0.001</td>
<td>91.2</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Farmer 9</td>
<td>1.22 (1.01)</td>
<td>-0.09</td>
<td>0.198</td>
<td>67.7</td>
<td>94.1</td>
<td>100</td>
</tr>
<tr>
<td>Farmer 10</td>
<td>1.34 (1.05)</td>
<td>0.35</td>
<td>0.678</td>
<td>50.0</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td><strong>Overall Mean (SD)</strong></td>
<td>0.00 (0.06)</td>
<td></td>
<td></td>
<td>68.3 (12.2)</td>
<td>98.0 (2.4)</td>
<td></td>
</tr>
</tbody>
</table>

Comparison of difference in means between VS and Farmer observers (P value)

| | 0.5 | 1.0 | 0.8 | 0.8 | 180 | 181 | 182 |
a) Bias between attempts

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

b) Exact agreement

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

c) Pairwise kappa

The mean $\kappa_W$ at intra-observer level was 0.81 for VS and 0.83 for farmers (P = 0.8).

Inter-observer agreement

a) Kappa between observers (Table 1)

The mean $\kappa_W$ at inter-observer level was 0.73 (SD 0.04) for VS and 0.72 (SD 0.04) for farmers (P = 0.8).

b) Kappa for locomotion scores (Table 2)

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

c) Median scores

The median score assigned to each video clip by VS was not significantly different to that assigned by farmers (P = 0.18) (Table 2).
Table 2: Inter-observer agreement for individual locomotion scores

<table>
<thead>
<tr>
<th>Locomotion score</th>
<th>Overall κ VS and Farmers§</th>
<th>Overall κ VS§</th>
<th>Overall κ Farmers§</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.37 Fair</td>
<td>0.30 Fair</td>
<td>0.43 Moderate</td>
</tr>
<tr>
<td>1</td>
<td>0.22 Fair</td>
<td>0.27 Fair</td>
<td>0.18 Slight</td>
</tr>
<tr>
<td>2</td>
<td>0.43 Moderate</td>
<td>0.47 Moderate</td>
<td>0.41 Moderate</td>
</tr>
<tr>
<td>3</td>
<td>0.62 Substantial</td>
<td>0.67 Substantial</td>
<td>0.58 Moderate</td>
</tr>
<tr>
<td><strong>Combined</strong></td>
<td><strong>0.40 Fair</strong></td>
<td><strong>0.42 Moderate</strong></td>
<td><strong>0.39 Fair</strong></td>
</tr>
</tbody>
</table>

§ 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
DAIRYCO (2009) Mobility Scoring. In Mobility scoring tool, DairyCo
In Farm Animal Welfare Council Reports