***Supplementary Table S4*** Comparison of assessment scores using the full model or simplification strategy model 1 (predicting gait scores from measures of hock burn on-farm) or 2 (substitution of on-farm measures by slaughterhouse scores), indicated as correlation between full and simplified model calculations (Spearman rank correlation; rsp) and distribution of flocks across classification groups according to Welfare Quality® (< 20, between 20 and 55, >55). The table shows estimates for agreement between methods (Est.), and the lower(Low) and upper (Upp) limits within classification groups for equality (% eq);specificity (%sp) and sensitivity (%se) of estimation. Empty cells indicate that there were no flocks with assessment scores in these classification groups.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Principle/criterion | Simplification strategy | | RSp |  | Estimates for agreement | | | | |  | | | | | | | | |
|  |  |  | |  | Class. < 20 | | | Class < 55 | | | | Class>55 | | | | | | | |
|  |  |  | |  | Est2 | Low2 | Upp2 | Est | Low | | Upp | | Est | Low | | Upp | |
| P2. Good housing | Model 2 | 0.99 | | %eq1 | 98.5 | 95.5 | 99.7 | 96.4 | 92.5 | | 98.6 | | 100 | | 97.8 | | 100 | |
|  |  |  | | %sp1 | 97.8 | 89.9 | 99.9 | 98.2 | 94.5 | | 99.7 | | 100 | | 97.8 | | 100 | |
|  |  |  | | %se1 | 98.9 | 95 | 99.9 | 87.5 | 70.8 | | 96.5 | |  | |  | |  | |
| P3. Good health | Model 1 | 0.79 | | %eq | 100 | 97.7 | 100 | 100 | 97.7 | | 100 | | 100 | | 97.7 | | 100 | |
|  |  |  | | %sp |  |  |  | 100 | 97.7 | | 100 | | 100 | | 97.7 | | 100 | |
|  |  |  | | %se | 100 | 97.7 | 100 |  |  | |  | |  | |  | |  | |
|  | Model 2 | 0.61 | | %eq | 100 | 97.7 | 100 | 100 | 97.7 | | 100 | | 100 | | 97.7 | | 100 | |
|  |  |  | | %sp |  |  |  | 100 | 97.7 | | 100 | | 100 | | 97.7 | | 100 | |
|  |  |  | | %se | 100 | 97.7 | 100 |  |  | |  | |  | |  | |  | |
| C3.Comfort around resting | Model 2 | 0.91 | | %eq | 100 | 97.9 | 100 | 97.2 | 93.6 | | 99 | | 100 | | 97.9 | | 100 | |
|  |  |  | | %sp | 100 | 36.8 | 100 | 100 | 97.6 | | 100 | | 100 | | 97.9 | | 100 | |
|  |  |  | | %se | 100 | 97.8 | 100 | 77.8 | 56.1 | | 92 | |  | |  | |  | |
| C6. Absence of | Model 1 | 0.81 | | %eq | 88.2 | 82.8 | 92.3 | 100 | 97.9 | | 100 | | 100 | | 97.9 | | 100 | |
| injury |  |  | | %sp | 94.1 | 88.6 | 97.4 | 100 | 97.9 | | 100 | | 100 | | 97.9 | | 100 | |
|  |  |  | | %se | 74.4 | 61.2 | 84.9 |  |  | |  | |  | |  | |  | |
|  | Model 2 | 0.68 | | %eq | 86.1 | 80.5 | 90.6 | 100 | 97.9 | | 100 | | 100 | | 97.9 | | 100 | |
|  |  |  | | %sp | 95 | 89.9 | 98 | 100 | 97.9 | | 100 | | 100 | | 97.9 | | 100 | |
|  |  |  | | %se | 65.1 | 51.5 | 77.1 |  |  | |  | |  | |  | |  | |

1 equality (% eq);specificity (%sp) and sensitivity (%se) of estimation

2 estimates for agreement between methods (Est), and the lower (Low) and upper (Upp) limits within classification groups