

Welfare Monitoring System Horses

Calculation of scores



LIVESTOCK RESEARCH
WAGENINGEN UR

Colophon

Welfare Monitoring System Horses – Calculation of scores – version 2.0

Report number 570

Publisher

Wageningen UR Livestock Research

Postbus 65, 8200 AB Lelystad

Copyright

© Wageningen UR Livestock Research, part
of Stichting Dienst Landbouwkundig
Onderzoek (DLO Foundation), March 2012

Reproduction of contents, either whole or in
part, permitted with due reference to the
source.

Liability

Wageningen UR Livestock Research does not
accept any liability for damages, if any,
arising from the use of the results of this
study or the application of the
recommendations.

Wageningen UR Livestock Research and
Central Veterinary Institute of Wageningen
UR, both part of Stichting Dienst
Landbouwkundig Onderzoek (DLO
Foundation), together with the Department of
Animal Sciences of Wageningen University
comprises the Animal Sciences Group of
Wageningen UR (University & Research
centre).

CONSTRUCTION OF WELFARE CRITERIA FOR HORSES ACCORDING TO THE WELFARE QUALITY® APPROACH

Introduction

In accordance with the Welfare Quality® (WQ) approach, an integration model has been constructed for the calculation of scores for welfare criteria for horses, based on expert opinion. A total of 10 experts were consulted during four 1-day sessions. Similar to the consultation of experts in previous projects that were part of WQ, experts were asked to assign scores to predefined situations and/or outcomes on virtual farms. These scores were then processed in agreement with the mathematical methodology that was developed within WQ (Botreau, 2008; Bonde et al, 2009). This report describes the methods proposed to calculate criterion-scores for horses.

Comprehensive descriptions of the welfare criteria, and of the various (animal- and/or environment-based) measures underlying each criterion are provided elsewhere (see Welfare Monitoring System Horses – assessment protocol version 2.0).

Below, for each criterion, the method for the construction of scores is described.

1. Criterion 1: Absence of prolonged hunger

MEASURES

Poor body condition

Classification: 0 = Very poor or poor, 1 = Moderate, 2 = Good.

The % of horses with a poor or very poor body condition and the % of horses with a moderate body condition are first combined in a weighted sum, and then transformed into an index that varies between 0 and 100 as follows:

$$I = 100 - [2(\% \text{ moderate}) + 3(\% \text{ very poor or poor})]/3$$

This index is converted into a score between 0 and 100 with the use of a spline function (Figure 1), as follows:

Coefficients:

	COEF	VALUE
a1		0.0000000000
b1		0.2728151601
c1		-0.0034101895
d1		0.0000717011
a2		-2247.6311658916
b2		84.5589821896
c2		-1.0569872582
d2		0.0044616055

Index = I

When $I \leq 80$ then Score = $b1 \times I + c1 \times I^2 + d1 \times I^3$

When $I \geq 80$ then Score = $a2 + b2 \times I + c2 \times I^2 + d2 \times I^3$

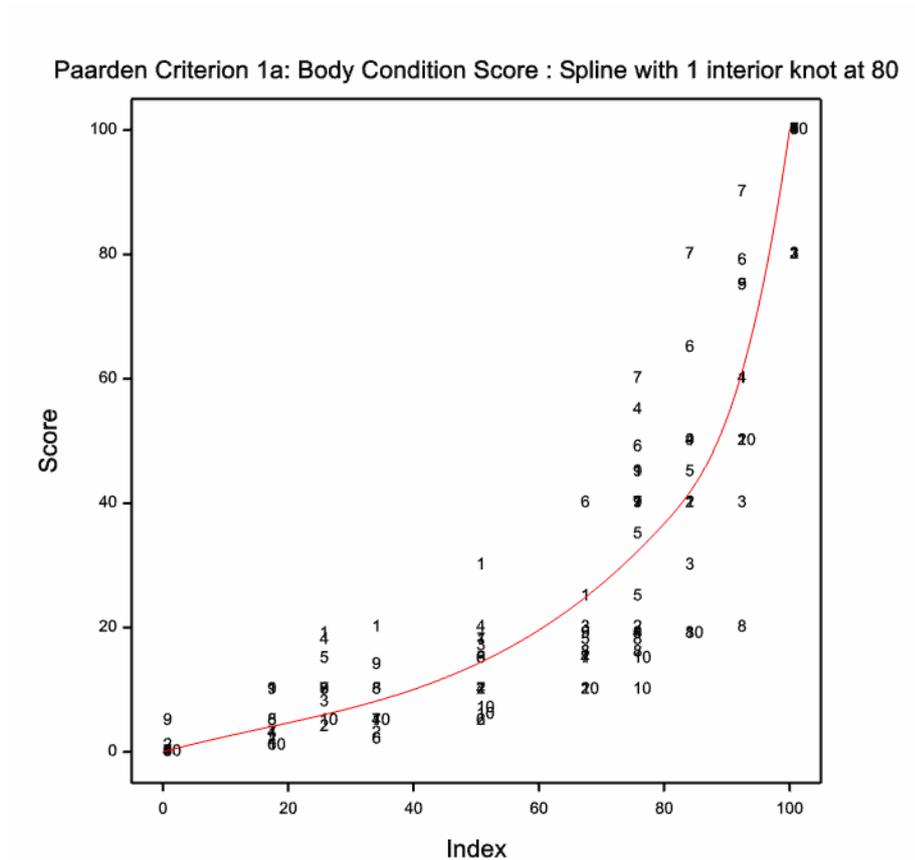


Figure 1. Calculation of the score for Poor body condition according to the % horses that are moderately poor and the % horse that are very poor or poor.

Wear pattern incisors: feed intake

Classification: 0 = normal, 1 = abnormal

The % of horses with abnormal wear pattern incisors is transformed into an index that varies between 0 and 100 as follows:

$$I = 100 - \% \text{ abnormal}$$

This index is converted into a score between 0 and 100 with the use of a spline function (Figure 2), as follows:

Coefficients:

	COEF	VALUE
a1		0.0000000000
b1		0.5693682727

c1	-0.0062657212
d1	0.0000820194
a2	-1516.8642499847
b2	57.4517777263
c2	-0.7172958403
d2	0.0030446449

Index = I

When $I \leq 80$ then Score = $b1 \times I + c1 \times I^2 + d1 \times I^3$
 When $I \geq 80$ then Score = $a2 + b2 \times I + c2 \times I^2 + d2 \times I^3$

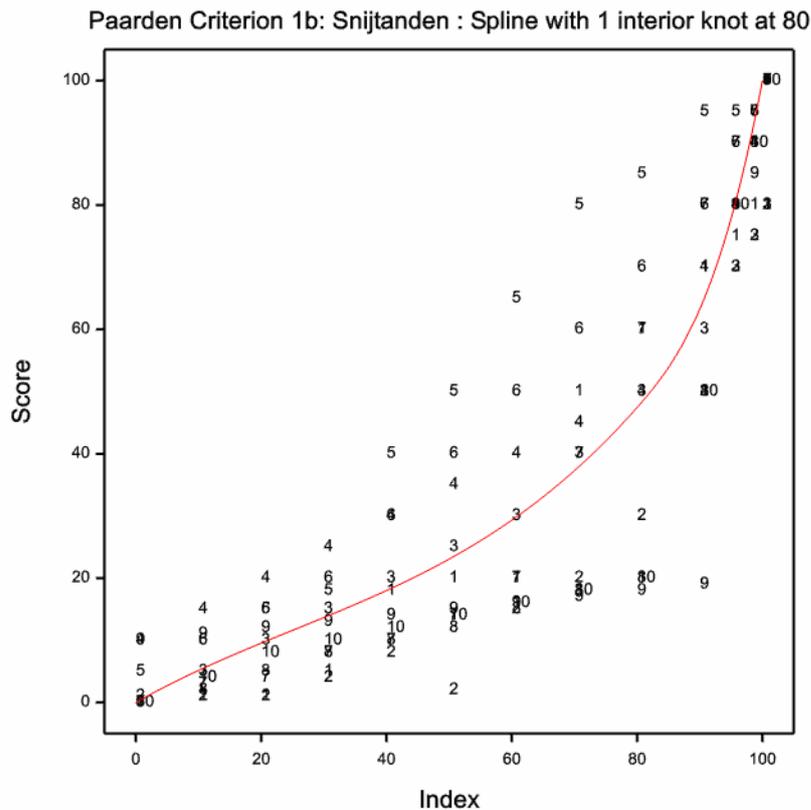


Figure 2. Calculation of the score for Abnormal wear pattern incisors.

Inspection of horse teeth

Classification: 0 = At least once a year, 1 = Less than once a year, 2 = no inspection.

The % of horses with less than once a year and the % of horses with no inspection of the teeth are first combined in a weighted sum, and then transformed into an index that varies between 0 and 100 as follows:

$$I = 100 - [3(\% \text{ less than once}) + 5(\% \text{ no inspection})]/5$$

This index is converted into a score between 0 and 100 with the use of a spline function (Figure 3), as follows:

Coefficients:

	COEF	VALUE
a1	0.0000000000	
b1	1.2336045441	
c1	-0.0154200568	
d1	0.0001352760	
a2	283.8976422471	
b2	-9.4125569877	
c2	0.1176569617	
d2	-0.0004192116	

Index = I

When $I \leq 80$ then Score = $b1 \times I + c1 \times I^2 + d1 \times I^3$
 When $I \geq 80$ then Score = $a2 + b2 \times I + c2 \times I^2 + d2 \times I^3$

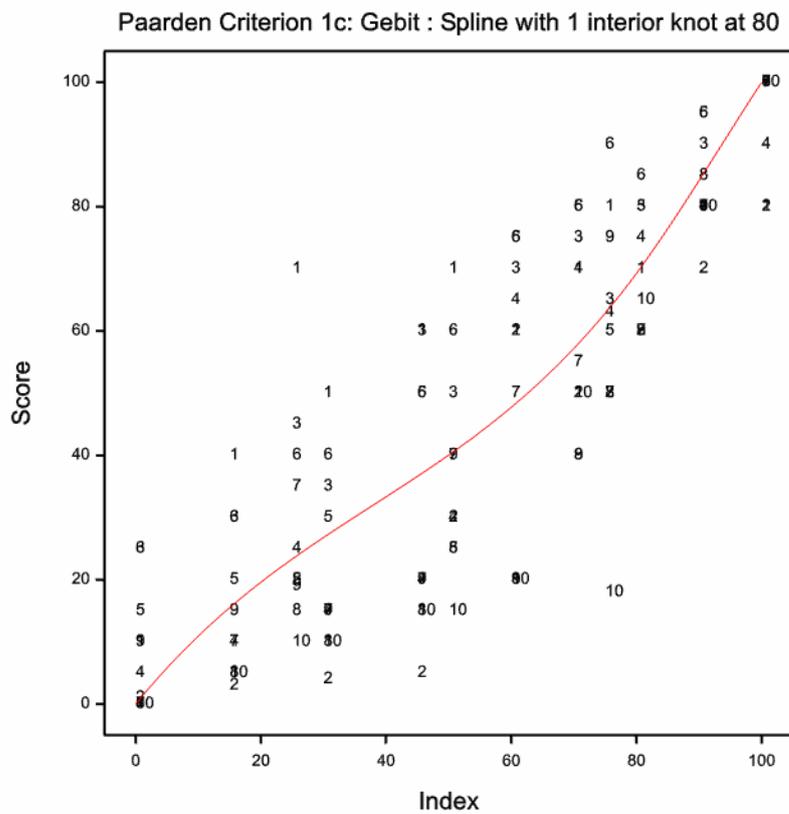


Figure 3. Calculation of the score for Inspection of horse teeth.

Feed intake

Different situations with regard to feed intake are represented in a decision tree (21 outcomes, A – U). Each individual horse in the sample receives a score according to the feed intake situation. The measure “Feed intake” at farm level is the average outcome score of all horses in the sample.

INTEGRATION OF MEASURES INTO CRITERION SCORE

For the integration of the four measures underlying Criterion 1 (see above), a Choquet integral was calculated, with 14 parameters (m):

Choquet integration

m1	0.3736	Poor body condition
m2	0.0278	Wear pattern incisors
m3	0.1303	Inspection teeth
m4	0.3995	Feed intake
m12	0.3333	
m13	0.3833	
m14	0.6767	
m23	0.1267	
m24	0.3033	
m34	0.3267	
m123	0.3561	
m124	0.7952	
m134	0.7744	
m234	0.3786	

2. Criterion 2: Absence of prolonged thirst

MEASURE

Water provision

Different situations with regard to water provision are represented in a decision tree (16 outcomes, A – P). Each individual horse in the sample receives a score according to the water provision situation. The score for Criterion 2 at farm level is the average outcome score of all horses in the sample.

3. Criterion 3: Comfort around resting

MEASURE

Comfort around resting

Different situations with regard to comfort around resting are represented in a decision tree (18 outcomes, A – R). Each individual horse in the sample receives a score according to the comfort around resting situation. The score for Criterion 3 at farm level is the average outcome score of all horses in the sample.

4. Criterion 4: Thermal comfort

MEASURE

Climate

Different situations with regard to climate are represented in a decision tree (16 outcomes, A – P). Each individual horse in the sample receives a score according to the climate situation. The score for Criterion 4 at farm level is the average outcome score of all horses in the sample.

5. Criterion 5: Ease of movement

MEASURE

Ease of movement

Different situations with regard to ease of movement are represented in a decision tree (22 outcomes, A – V). Each individual horse in the sample receives a score according to the ease of movement situation. The score for Criterion 5 at farm level is the average outcome score of all horses in the sample.

6. Criterion 6: Absence of injuries

MEASURES

Patches of white hairs

Classification: 0 = Normal, 1 = Abnormal.

During the welfare assessment on-farm, the presence of white hairs is recorded in each horse at five locations on the body. Following a quantitative rule which takes the total presence of white hairs into account, an individual horse receives one of two final scores with regard to the presence of white hairs: 0 = Normal, or 1 = Abnormal.

Experts were asked to define an Alarm threshold for the % horse with white hairs (i.e. % Abnormal).

A Warning threshold for this measure was defined as (Alarm threshold)/2.

Wounds

Classification: 0 = Normal, 1 = Moderately wounded, 2 = Severely wounded

During the welfare assessment on-farm, the presence of wounds is recorded in each horse at five locations on the body. Following a quantitative rule – based on a decision tree – which takes the total presence of wounds into account, an individual horse receives one of three final scores with regard to the presence of wounds: 0 = Normal, or 1 = Moderately wounded, 2 = Severely wounded.

The % of moderately wounded horses and the % of severely wounded horses are first combined in a weighted sum, and then transformed into an index that varies between 0 and 100 as follows:

$$I = 100 - [1(\% \text{ moderate}) + 3(\% \text{ severe})]/3$$

This index is converted into a score between 0 and 100 with the use of a spline function (Figure 4), as follows:

Coefficients:

	COEF	VALUE
a1		0.0000000000
b1		0.1964022151
c1		0.0018343760
d1		0.0000129238
a2		-8932.9659032198
b2		315.4775403427
c2		-3.7073553594
d2		0.0145587655

Index = I

When $I \leq 85$ then Score = $b1 \times I + c1 \times I^2 + d1 \times I^3$

When $I \geq 85$ then Score = $a2 + b2 \times I + c2 \times I^2 + d2 \times I^3$

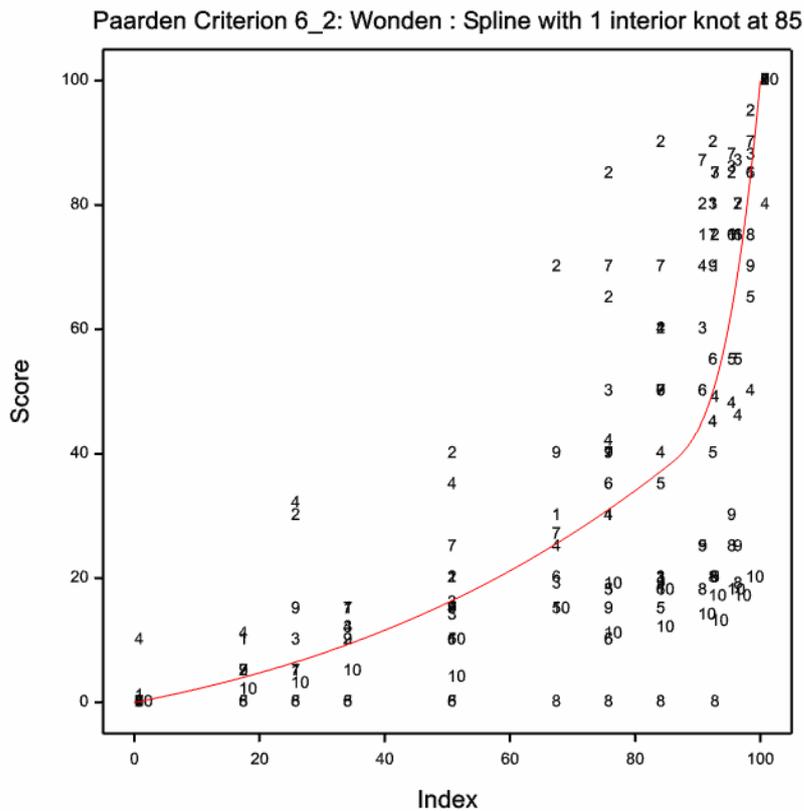


Figure 4. Calculation of the score for Wounds according to the % horses that are moderately wounded and the % horses that are severely wounded. Using the curve presented in Figure 4, the index for Wounds (x-axis) corresponding with a score of 20 (y-axis) is used as the Alarm threshold, and the index for Wounds corresponding with a score of 40 is used as the Warning threshold.

Lameness

Classification: 0 = Normal, 1 = Moderately lame (irregular, stiff), 2 = Severely lame (clear evidence of lameness)

The % of moderately lame horses and the % of severely lame horses are first combined in a weighted sum, and then transformed into an index that varies between 0 and 100 as follows:

$$I = 100 - [2(\% \text{ moderate}) + 5(\% \text{ severe})]/5$$

This index is converted into a score between 0 and 100 with the use of a spline function (Figure 5), as follows:

Coefficients:

	COEF	VALUE
a1		0.0000000000
b1		0.1263848002
c1		0.0013530760
d1		0.0000064387
a2		-12262.8733399423

Hoof condition

Classification: 0 = No signs of neglect ion, 1 = Signs of neglect ion.

Experts were asked to define an Alarm threshold for the % horse with signs of neglect ion of the hooves (i.e. % Abnormal).

A Warning threshold for this measure was defined as (Alarm threshold)/2.

Swollen legs

Classification: 0 = Normal, 1 = Swollen legs

Experts were asked to define an Alarm threshold for the % horse with swollen legs.

A Warning threshold for this measure was defined as (Alarm threshold)/2.

Safety

Safety is assessed at two levels: (i) at the level of the farm, and (ii) at the level of the individual horse.

Safety at farm level (public area)

Classification: 0 = Safe (no or minor risk), 1 = Unsafe (moderate or large risk).

If the situation at farm level is classified as unsafe, the measure "Safety" is automatically considered to have exceeded the Alarm level. If the situation at farm level is classified as safe, observations with regard to the safety at horse level are taken into consideration (see below). These latter observations will then determine whether or not the measure "Safety" has exceeded either the Warning or the Alarm level.

Safety at horse level (horse area).

Classification: 0 = Safe, 1 = Unsafe.

Experts were asked to define an Alarm threshold for the % horses exposed to moderate or large risks.

A warning threshold was defined as (Alarm threshold)/2

INTEGRATION OF MEASURES INTO CRITERION SCORE

The 7 measures underlying Criterion 6 (see above) are integrated into a Criterion score based on the total number of "Alarms" and "Warnings" per farm. The total number of Alarms and Warnings per farm were evaluated by the experts in terms of a score between 0 and 100.

Establishing the total number of Alarms and Warnings per farm occurs in two steps:

First, the measures "Wounds" and "Swollen legs" are combined in the sense that a single Alarm or Warning is attributed based on the outcome of both Wounds and Swollen legs. In this respect, the following rule is used (see Table 1).

Table 1. Single Alarm, Warning or no Alarm or Warning, based on the combination of the measures Wounds and Swollen legs.

Wounds	Swollen legs		
	A	W	O
A	A	A	W
W	A	W	W
O	A	W	O

Secondly, the total number of Alarms and Warnings per farm is counted. An individual farm may have a maximum of 6 Alarms. An individual farm without any problems in terms of Criterion 6 will have no Alarms and no Warnings. In between are farms with differential numbers of Alarms and Warning. Differential outcomes in terms number of Alarms and Warnings were scored by experts.

An index is calculated as :

$$I = \left(\frac{100}{N_{tot}} \times \left(N_{tot} - \frac{\sum_{k=0}^2 w_k N_k}{w_2} \right) \right)$$

with N_{tot} = number of areas = 6

N_k number of problems according to intensity

k= 0 no problems

k= 1 warnings

k= 2 alarms

w_k weight of problems

w_0 = 0

w_1 = 1

w_2 = 2

Then the index is transformed into a score according to I-spline functions as follows (Figure 6).

Paarden Criterion 6: Absence of injuries : Spline with 1 interior knot at 40

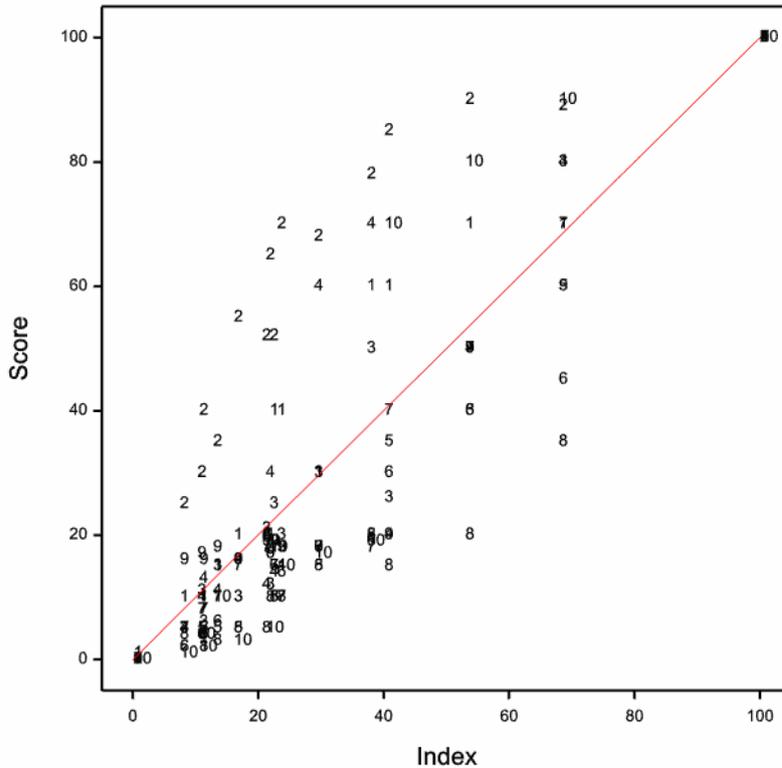


Figure 6. Calculation of scores for absence of injuries according to the proportion of areas for which symptoms are above the warning or the alarm threshold

7. Criterion 7: Absence of disease

MEASURES

In total, 9 health measures are recorded in every individual horse in the sample. These 9 measures are distributed across 5 'Areas', see Table 2.

Table 2: Distribution of health measures across 5 'Areas'

Area	Measure
1. Respiratory system	Abnormal breathing
	Coughing
	Nasal discharge
2. Skin and coat	Generalised skin problem
	Coat condition
	Rubbed and broken hairs in mane and tail
	Itchiness
3. Eyes	Ocular discharge
4. Fat body condition	Fat body condition
5. Skin irritation lower legs	Skin irritation lower legs

Ocular discharge

Classification: 0 = Normal, 1 = Dirty, 2 = Dirty – visible discharge

The % of horses with dirty eyes and the % of horses with dirty eyes as well as visible discharge are first combined in a weighted sum, and then transformed into an index that varies between 0 and 100 as follows:

$$I = 100 - [1(\% \text{ dirty}) + 2(\% \text{ dirty} - \text{visible discharge})]/2$$

This index is converted into a score between 0 and 100 with the use of a spline function (Figure 7), as follows:

Coefficients:

	COEF	VALUE
a1	0.0000000000	
b1	0.4330327264	
c1	-0.0050945027	
d1	0.0000516509	
a2	-10188.2590090594	
b2	360.0186727302	
c2	-4.2355141028	
d2	0.0166415328	

Index = I

When $I \leq 85$ then Score = $b1 \times I + c1 \times I^2 + d1 \times I^3$

When $I \geq 85$ then Score = $a2 + b2 \times I + c2 \times I^2 + d2 \times I^3$

Using the curve presented in Figure 7, the index for Ocular discharge (x-axis) corresponding with a score of 20 (y-axis) is used as the Alarm threshold, and the index for Ocular discharge corresponding with a score of 40 is used as the Warning threshold.

Index = I

When $I \leq 85$ then Score = $b_1 \times I + c_1 \times I^2 + d_1 \times I^3$

When $I \geq 85$ then Score = $a_2 + b_2 \times I + c_2 \times I^2 + d_2 \times I^3$

Using the curve presented in Figure 8, the index for Fat body condition (x-axis) corresponding with a score of 20 (y-axis) is used as the Alarm threshold, and the index for Fat body condition corresponding with a score of 40 is used as the Warning threshold.

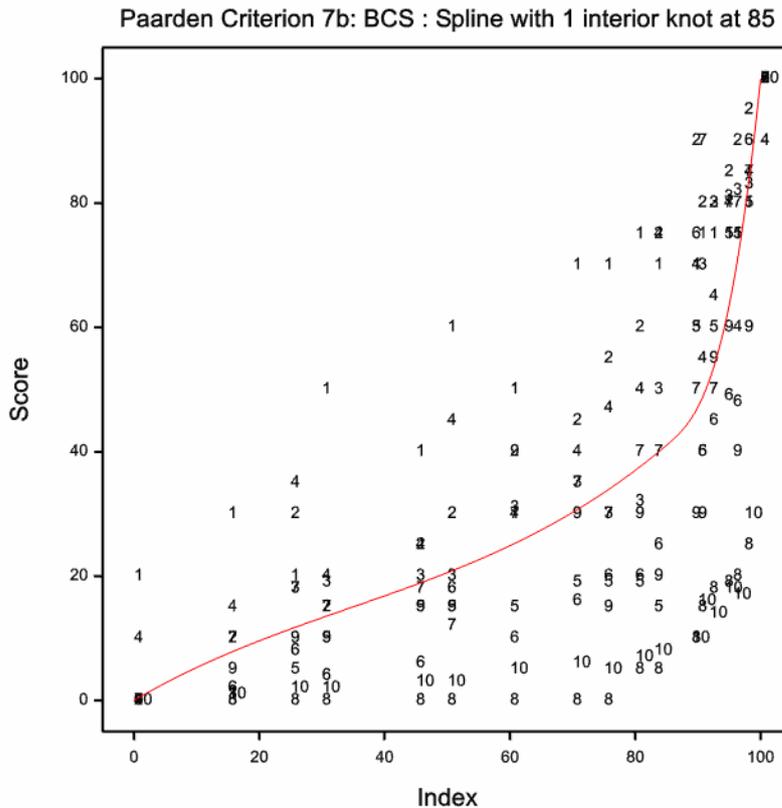


Figure 8. Calculation of the score for Fat body condition according to the % of fat horses and the % of very fat horses.

Skin irritation lower legs

Classification: 0 = Normal, 1 = Moderate, 2 = Severe

The % of horses with moderate skin irritation and the % of horses with severe skin irritation at the lower legs are first combined in a weighted sum, and then transformed into an index that varies between 0 and 100 as follows:

$$I = 100 - [4(\% \text{ moderate}) + 15(\% \text{ severe})]/15$$

This index is converted into a score between 0 and 100 with the use of a spline function (Figure 9), as follows:

Coefficients:

	COEF	VALUE
a1	0.0000000000	
b1	0.0671880259	
c1	0.0045339365	
d1	-0.0000355281	
a2	-15188.4668411841	
b2	536.1306295852	
c2	-6.3020937565	
d2	0.0246963415	

Index = I

When $I \leq 85$ then Score = $b1 \times I + c1 \times I^2 + d1 \times I^3$

When $I \geq 85$ then Score = $a2 + b2 \times I + c2 \times I^2 + d2 \times I^3$

Using the curve presented in Figure 9, the index for Skin irritation at the lower legs (x-axis) corresponding with a score of 20 (y-axis) is used as the Alarm threshold, and the index for Skin irritation at the lower legs corresponding with a score of 40 is used as the Warning threshold

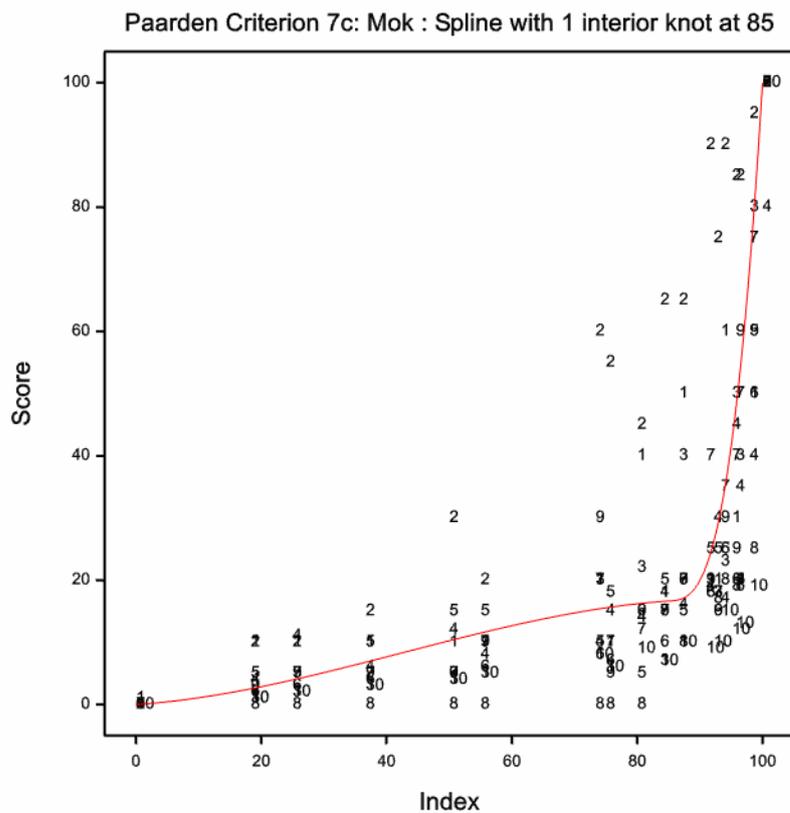


Figure 9. Calculation of the score for Skin irritation lower legs according to the % of horses with moderate and the % of horses with severe irritation.

Abnormal breathing

Classification: 0 = Normal, 1 = Abnormal.

Experts were asked to define an Alarm threshold for the % horses with abnormal breathing.

A Warning threshold for this measure was defined as (Alarm threshold)/2.

Coughing

Classification: 0 = Normal, 1 = Coughing

Experts were asked to define an Alarm threshold for the % horses exhibiting coughing.
A Warning threshold for this measure was defined as (Alarm threshold)/2.

Nasal discharge

Classification: 0 = Normal, 1 = Nasal discharge

Experts were asked to define an Alarm threshold for the % horses exhibiting nasal discharge.

A Warning threshold for this measure was defined as (Alarm threshold)/2.

Generalised skin problem

Classification: 0 = Normal, 1 = Skin problem

Experts were asked to define an Alarm threshold for the % horses with a generalised skin problem.

A Warning threshold for this measure was defined as (Alarm threshold)/2.

Coat condition

Classification: 0 = Normal, 1 = Abnormal

Experts were asked to define an Alarm threshold for the % horses with an abnormal coat condition.

A Warning threshold for this measure was defined as (Alarm threshold)/2.

Itchiness

Classification: 0 = Normal, 1 = Itching

Experts were asked to define an Alarm threshold for the % horses with signs of itching.
A Warning threshold for this measure was defined as (Alarm threshold)/2.

Rubbed and broken hairs in mane and tail

Classification: 0 = Normal, 1 = Moderate, 2 = Severe

The % of horses with moderately rubbed and broken hairs in mane and tail the % of horses with severely rubbed and broken hairs in mane and tail are first combined in a weighted sum, and then transformed into an index that varies between 0 and 100 as follows:

$$I = 100 - [1(\% \text{ moderate}) + 5(\% \text{ severe})]/5$$

This index is converted into a score between 0 and 100 with the use of a spline function (Figure 10), as follows:

Coefficients:

	COEF	VALUE
a1	0.0000000000	
b1	0.1776870082	
c1	0.0032104787	
d1	-0.0000241541	
a2	-13516.3520637968	
b2	477.2253097941	
c2	-5.6091134244	
d2	0.0219849553	

Index = I

When $I \leq 85$ then Score = $b1 \times I + c1 \times I^2 + d1 \times I^3$

When $I \geq 85$ then Score = $a2 + b2 \times I + c2 \times I^2 + d2 \times I^3$

Using the curve presented in Figure 10, the index for Rubbed and broken hairs in mane and tail (x-axis) corresponding with a score of 20 (y-axis) is used as the Alarm threshold, and the index for Rubbed and broken hairs in mane and tail corresponding with a score of 40 is used as the Warning threshold.

Paarden Criterion 7d: Staarten en manen : Spline with 1 interior knot at 85

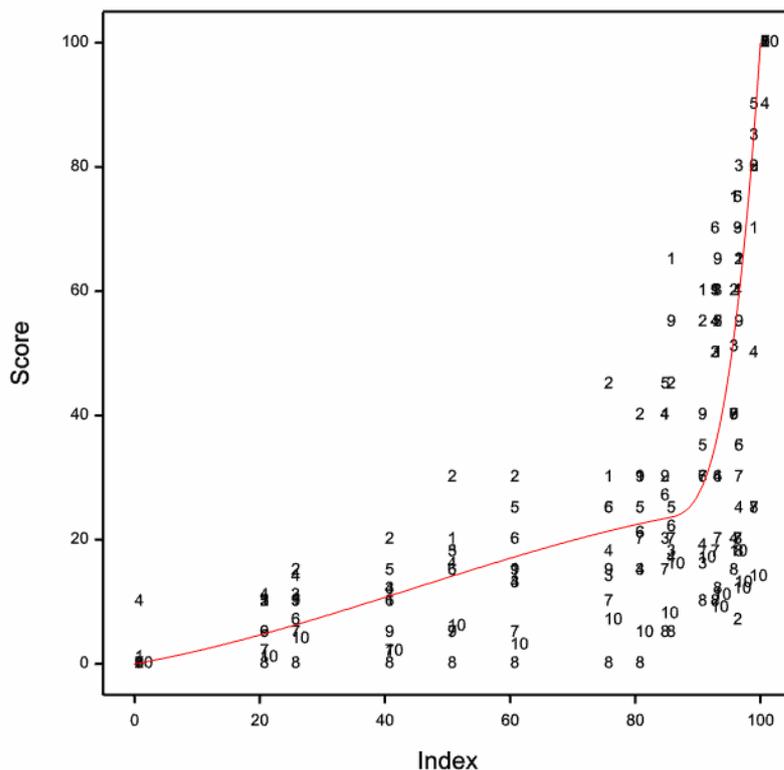


Figure 10. Calculation of the score for Rubbed and broken hairs in mane and tail according to the % of horses with moderately rubbed and broken hairs in mane and tail the % of horses with severely rubbed and broken hairs in mane and tail % of horses with moderate and the % of horses with severe irritation

INTEGRATION OF MEASURES INTO CRITERION SCORE

The 10 measures underlying Criterion 7 (see above) are integrated into a Criterion score based on the total number of “Alarms” and “Warnings” per farm. Alarms and Warnings are assigned to ‘Areas’ (see Table 2). As soon as the prevalence of at least one of the measures underlying the same Area is beyond the Warning threshold, and the prevalence of the other measures is lower than the Alarm threshold, a Warning is assigned to that particular Area. As soon as the prevalence of at least one of the measures underlying the same Area is beyond the Alarm threshold, an Alarm is assigned to that particular Area. The total number of Alarms and Warnings per farm is counted. An individual farm may have a maximum of 5 Alarms. An individual farm without any problems in terms of Criterion 7 will have no Alarms and no Warnings. In between are farms with differential numbers of Alarms and Warning. Differential outcomes in terms number of Alarms and Warnings were scored by experts.

An index is calculated as :

$$I = \left(\frac{100}{N_{tot}} \times \left(N_{tot} - \frac{\sum_{k=0}^2 w_k N_k}{w_2} \right) \right)$$

with N_{tot} = number of areas = 5

N_k number of problems according to intensity

k= 0 no problems

k= 1 warnings

k= 2 alarms

w_k weight of problems

$w_0= 0$

$w_1= 1$

$w_2= 2$

Paarden Criterion 7: Absence of disease : Spline with 1 interior knot at 40

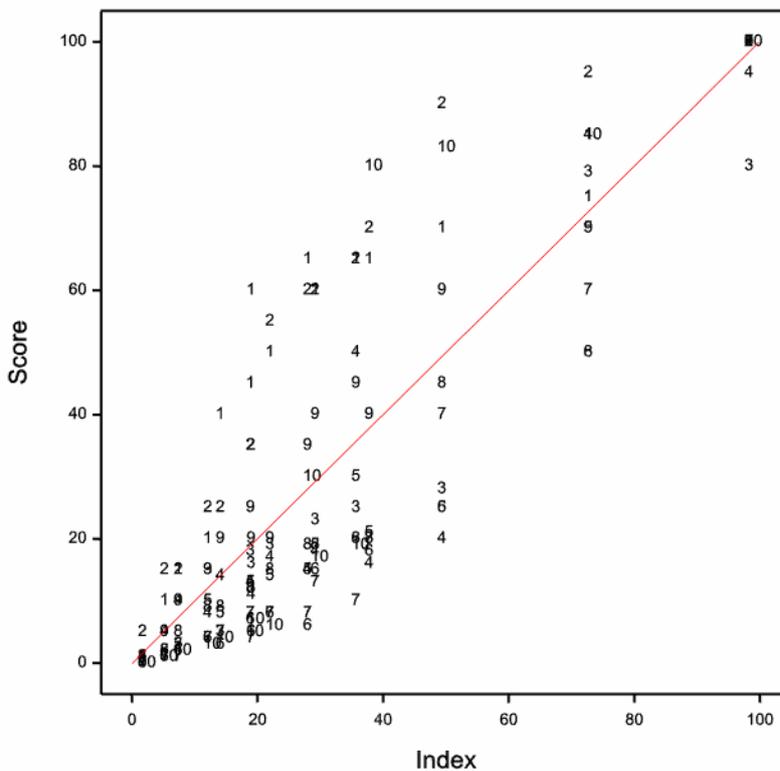


Figure 11. Calculation of scores for absence of disease according to the proportion of areas for which symptoms are above the warning or the alarm threshold

Then the index is transformed into a score according to I-spline functions as follows (Figure 11).

Coefficients:

	COEF	VALUE
a1		0.0000000000
b1		0.9933554026

c1	0.0005597151
d1	-0.0000104544
a2	0.1414784781
b2	0.9943768428
c2	0.0000412310
d2	0.0000001091

Index = I

When $I \leq 40$ then Score = $b1 \times I + c1 \times I^2 + d1 \times I^3$
 When $I \geq 40$ then Score = $a2 + b2 \times I + c2 \times I^2 + d2 \times I^3$

Criterion 8: Absence of discomfort caused by use

MEASURES

Back muscles

Classification: 0 = Normal, 1 = Tense, 2 = Very tense

The % of horses with tense and the % of horses with very tense back muscles are first combined in a weighted sum, and then transformed into an index that varies between 0 and 100 as follows:

$$I = 100 - [1(\% \text{ tense}) + 2(\% \text{ very tense})]/2$$

This index is converted into a score between 0 and 100 with the use of a spline function (Figure 12), as follows:

Coefficientse:

	COEF	VALUE
a1	0.0000000000	
b1	0.2204026222	
c1	-0.0025929720	
d1	0.0000536342	
a2	-9144.5793485763	
b2	322.9702904655	
c2	-3.7996507882	
d2	0.0149440582	

Index = I

When $I \leq 85$ then Score = $b1 \times I + c1 \times I^2 + d1 \times I^3$
 When $I \geq 85$ then Score = $a2 + b2 \times I + c2 \times I^2 + d2 \times I^3$

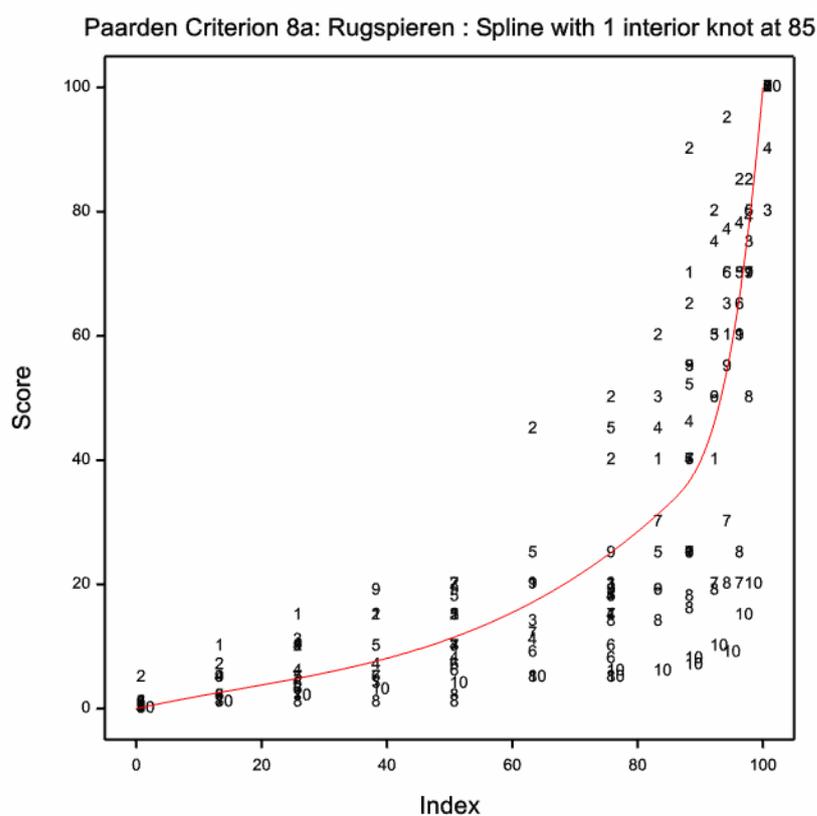


Figure 12. Calculation of the score for Back muscles according to the % of horses with tense and the % of horses with very tense back muscles.

Mouth corners

Classification: 0 = normal, 1 = abnormal

The % of horses with abnormal mouth corners is transformed into an index that varies between 0 and 100 as follows:

$$I = 100 - \% \text{ abnormal}$$

This index is converted into a score between 0 and 100 with the use of a spline function (Figure 13), as follows:

Coefficients:

	COEF	VALUE
a1	0.0000000000	
b1	0.2751174744	
c1	-0.0032366762	
d1	0.0000536508	
a2	-9317.2611397861	
b2	329.1196314990	
c2	-3.8719956999	
d2	0.0152252550	

Index = I

When $I \leq 85$ then Score = $b1 \times I + c1 \times I^2 + d1 \times I^3$
 When $I \geq 85$ then Score = $a2 + b2 \times I + c2 \times I^2 + d2 \times I^3$

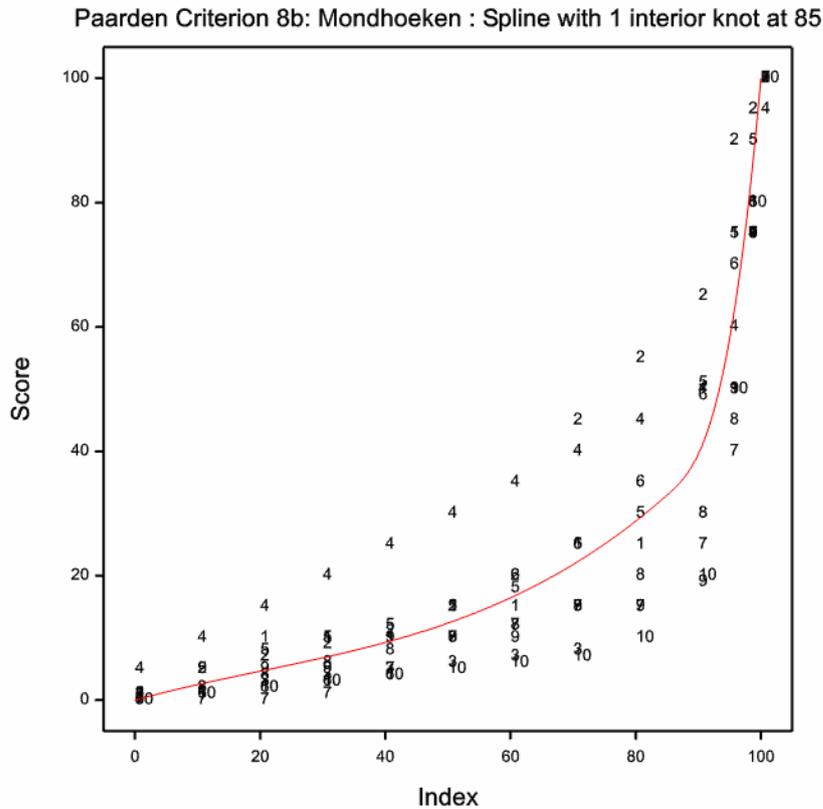


Figure 13. Calculation of the score for Mouth corners according to the % of horses with abnormal mouth corners.

Bars - fresh

Classification: 0 = normal, 1 = abnormal

The % of horses with abnormal bars (fresh wounds) is transformed into an index that varies between 0 and 100 as follows:

$$I = 100 - \% \text{ abnormal}$$

This index is converted into a score between 0 and 100 with the use of a spline function (Figure 14), as follows:

Coefficients:

	COEF	VALUE
a1		0.0000000000
b1		0.1600480990
c1		0.0023575186

d1	-0.0000189577
a2	-14443.7958118913
b2	509.9410817423
c2	-5.9950664613
d2	0.0235003523

Index = I

When $I \leq 85$ then Score = $b1 \times I + c1 \times I^2 + d1 \times I^3$

When $I \geq 85$ then Score = $a2 + b2 \times I + c2 \times I^2 + d2 \times I^3$

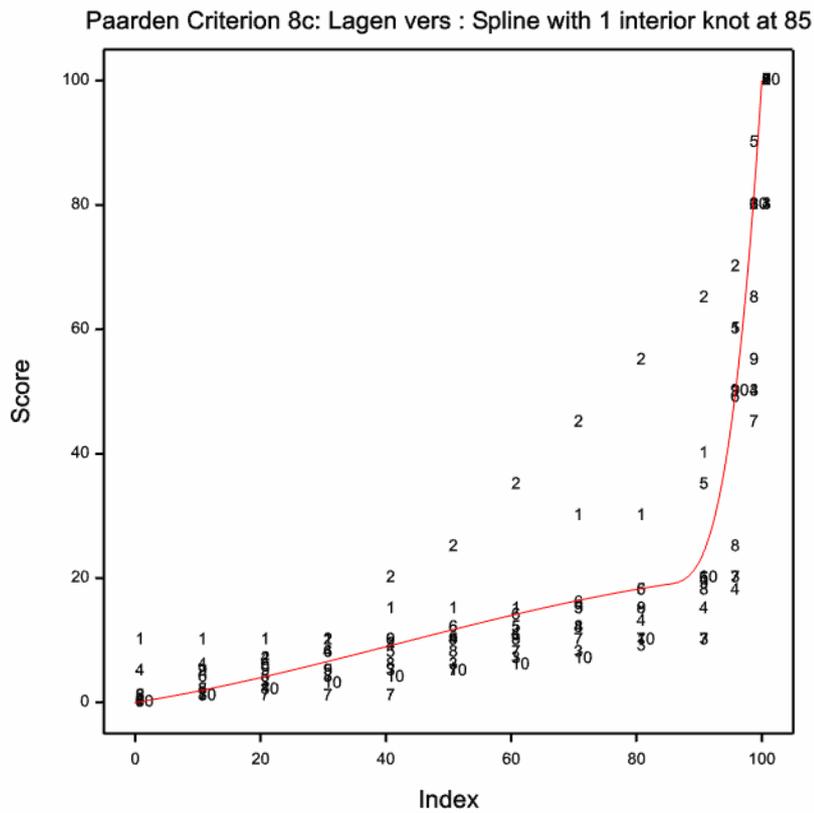


Figure 14. Calculation of the score for Bars – fresh according to the % of horses with abnormal bars (fresh wounds).

Bars – old

Classification: 0 = normal, 1 = abnormal

The % of horses with abnormal bars (old wounds) is transformed into an index that varies between 0 and 100 as follows:

$$I = 100 - \% \text{ abnormal}$$

This index is converted into a score between 0 and 100 with the use of a spline function (Figure 15), as follows:

Coefficients:

	COEF	VALUE
a1	0.0000000000	
b1	1.0087708291	
c1	-0.0113331575	
d1	0.0000914460	
a2	-3822.7669051068	
b2	135.9299570126	
c2	-1.5986412445	
d2	0.0063161836	

Index = I

When $I \leq 85$ then Score = $b1 \times I + c1 \times I^2 + d1 \times I^3$
 When $I \geq 85$ then Score = $a2 + b2 \times I + c2 \times I^2 + d2 \times I^3$

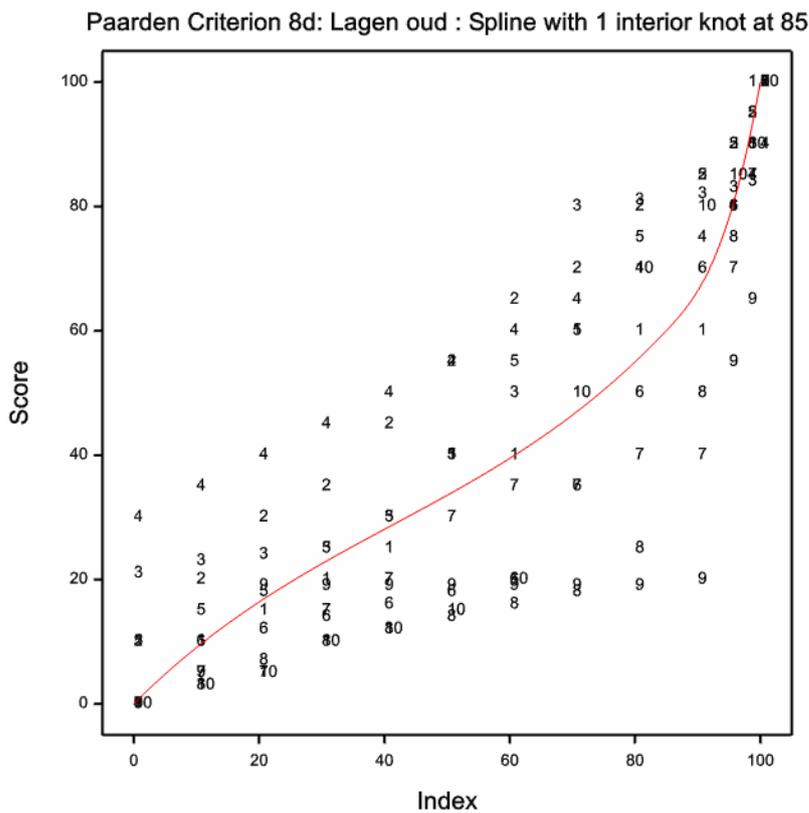


Figure 15. Calculation of the score for Bars – old according to the % of horses with abnormal bars (fresh wounds).

INTEGRATION OF MEASURES INTO CRITERION SCORE

For the integration of the four measures underlying Criterion 8 (see above), a Choquet integral was calculated, with 14 parameters (m):

Choquet integration

m1	0.8267	Back muscles
m2	0.6222	Mouth corners
m3	0.9000	Bars – fresh
m4	0.4044	Bars – old
m12	0.3533	
m13	0.5333	
m14	0.2033	
m23	0.4567	
m24	0.1167	
m34	0.1500	
m123	0.1000	
m124	0.1000	
m134	0.1000	
m234	0.1000	

Criterion 9: Expression of social behaviour

Different situations with regard to the expression of social behaviour are represented in a decision tree (5 outcomes, A – E). These situations are:

- Full social interaction
- Nibble and groom
- Sniff (through grid,, half open door)
- Visual contact
- No visual or physical contact

Experts were asked to assign a score to each of these situations, on the assumption that a horse is exposed to the situation for 24 hours.

Calculation of the scores for criterion 9 involves 2 steps:

For each individual horse the number of hours (with an accuracy of 1 hour), out of 24 hours, that the animal is exposed to each of the 5 situations is estimated.

Next, for each horse an individual score is calculated using the following formula:

$$\text{Score} = [(a.\text{Score}1) + (b.\text{Score}2) + (c.\text{Score}3) + (d.\text{Score}4) + (e.\text{Score}5)]/24$$

Where:

a = estimated number of hours in situation 1, b = estimated number of hours in situation 2, c = estimated number of hours in situation 3, d = estimated number of hours in situation 4, e = estimated number of hours in situation 5.

Score 1 = average score assigned by experts to situation 1, Score 2 = average score assigned by experts to situation 2, Score 3 = average score assigned by experts to situation 3, Score 4 = average score assigned by experts to situation 4, Score 5 = average score assigned by experts to situation 5

The score for Criterion 9 at farm level is the average of all individual scores of the horses in the sample.

Criterion 10: Expression of other behaviours

MEASURES

Wear pattern incisors: crib biting

Classification: 0 = normal, 1 = abnormal

The % of horses with abnormal wear pattern incisors indicative of crib biting is transformed into an index that varies between 0 and 100 as follows:

$$I = 100 - \% \text{ abnormal}$$

This index is converted into a score between 0 and 100 with the use of a spline function (Figure 16), as follows:

Coefficients:

	COEF	VALUE
a1		0.0000000000
b1		0.3077518710
c1		-0.0036206102
d1		0.0000298551
a2		-13751.9847059297
b2		485.6719533667
c2		-5.7137880776
d2		0.0224226701

Index = I

When $I \leq 85$ then Score = $b1 \times I + c1 \times I^2 + d1 \times I^3$

When $I \geq 85$ then Score = $a2 + b2 \times I + c2 \times I^2 + d2 \times I^3$

Paarden Criterion 10a: Snijtanden - kribbebijten : Spline with 1 interior knot at 85

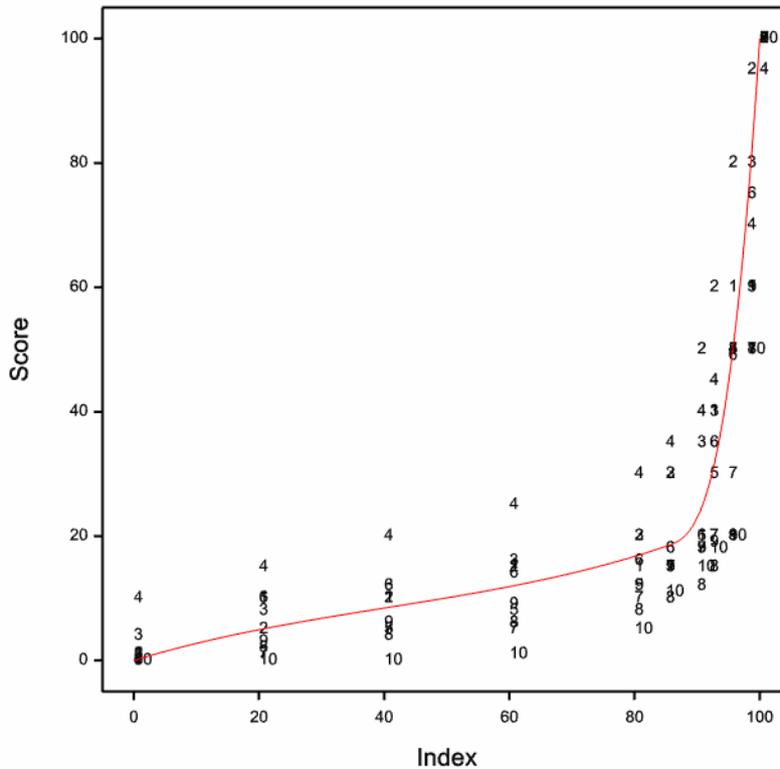


Figure 16. Calculation of the score Wear pattern incisors: crib biting according to the % of horses with abnormal wear pattern of the incisors indicative of crib biting.

Abnormal behaviours

Classification: 0 = normal, 1 = abnormal

The % of horses with abnormal wear pattern incisors indicative of crib biting is transformed into an index that varies between 0 and 100 as follows:

$$I = 100 - \% \text{ abnormal}$$

This index is converted into a score between 0 and 100 with the use of a spline function (Figure 17), as follows:

Coefficients:

	COEF	VALUE
a1	0.000000000	
b1	0.1289104971	
c1	-0.0015165941	
d1	0.0000251550	
a2	-14032.9520153890	
b2	495.4096060294	
c2	-5.8283487056	
d2	0.0228754785	

Index = I

When $I \leq 85$ then Score = $b_1 \times I + c_1 \times I^2 + d_1 \times I^3$
 When $I \geq 85$ then Score = $a_2 + b_2 \times I + c_2 \times I^2 + d_2 \times I^3$

Paarden Criterion 10b: Afwijkend gedrag : Spline with 1 interior knot at 85

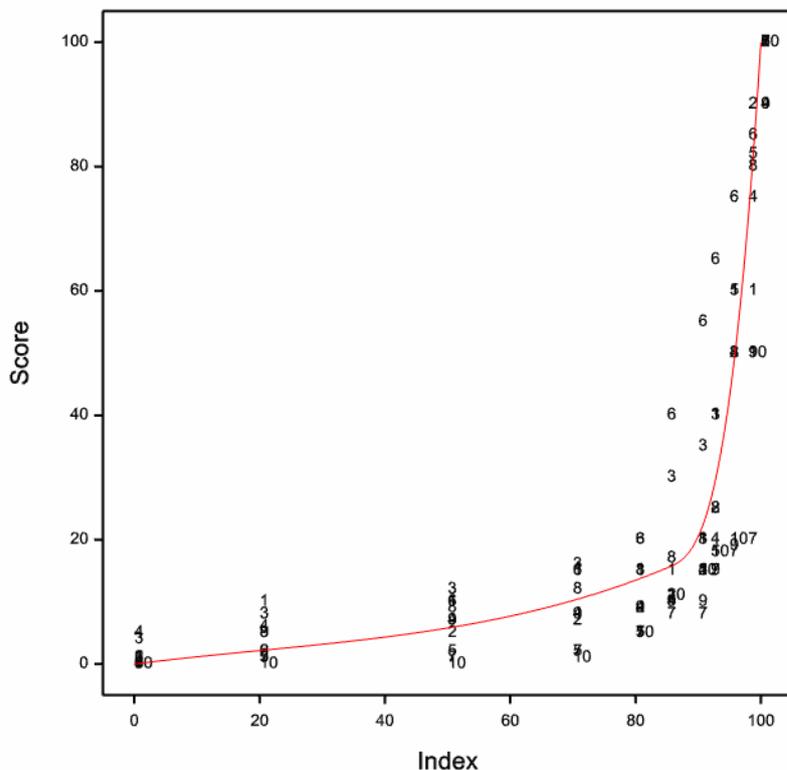


Figure 17. Calculation of the score Abnormal behaviours according to the % of horses exhibiting abnormal behaviours.

INTEGRATION OF MEASURES INTO CRITERION SCORE

For the integration of the four measures underlying Criterion 10 (see above), a Choquet integral was calculated, with 2 parameters (m):

Choquet integration

m1	0.3086	Wear pattern incisors: crib biting
m2	0.2655	Abnormal behaviours

Criterion 11: Good human animal relationship

No valid and feasible measure available.

Criterion 12: Positive emotional state

MEASURE

Visual horizon

Classification: 0 = Fully possible, 1 = Partially possible, 2 = Not possible

The % of horses with partial and the % of horses with full possibilities to broaden its visual horizon are first combined in a weighted sum, and then transformed into an index that varies between 0 and 100 as follows:

$$I = 100 - [2(\% \text{ partially possible}) + 2(\% \text{ fully possible})]/2$$

This index is converted into a score between 0 and 100 with the use of a spline function (Figure 18), as follows:

Coefficients:

	COEF	VALUE
a1	0.0000000000	
b1	1.2332692988	
c1	-0.0145090506	
d1	0.0001094942	
a2	-2232.5734364209	
b2	80.0299804888	
c2	-0.9415292007	
d2	0.0037448674	

Index = I

When $I \leq 85$ then Score = $b1 \times I + c1 \times I^2 + d1 \times I^3$

When $I \geq 85$ then Score = $a2 + b2 \times I + c2 \times I^2 + d2 \times I^3$

Paarden Criterion 12: Positive emotional state : Spline with 1 interior knot at 85

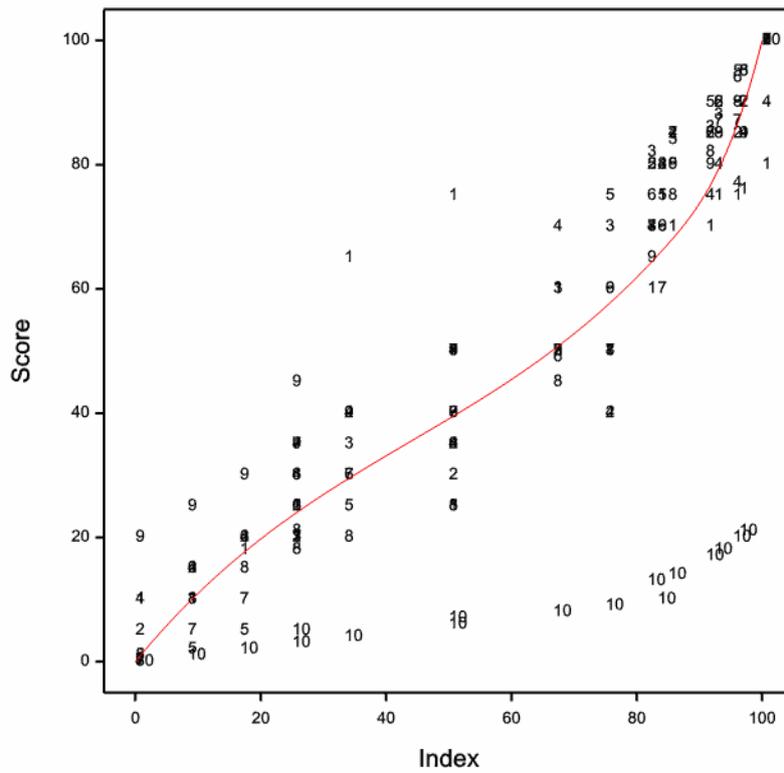


Figure 18. Calculation of the score for Visual horizon according to the % of horses with partial and the % of horses with full possibilities to broaden its visual horizon.

Correlation of Animal Based measures

In an attempt to simplify the horse welfare assessment at the farm correlation analysis has been executed on animal based measures collected during a national experiment with 150 farms and almost 3000 horses. This data has been collected between January and May 2010 with the first version of the Welfare Monitoring System assessment protocol for horses (version 1.0, Dutch).

A Pearson correlation analysis on farm level revealed several significant correlations between the animal based measures. However, none of the significant correlations has a biological logical reason for omitting one or the other. For example: significant correlation between “lameness” and “sensitivity of back muscles” (0.388) can be expected but by leaving for instance “lameness” out, one would miss a serious indication for a welfare risk. It was concluded that a data driven simplifying of the assessment protocol for horses (version 1.0) is not feasible.

Table 3. Pearson correlation animal based measures on farm level.

Measures	Back muscles	Lameness	Low BCS	High BCS	HAT avoid	Hoof condition	Skin irregularities	NOT avoid	Rubbed mane & tail	Stereotypies	White hairs	Wounds (cuts)	Bars	Mouth corners	Nasal discharge	Incisors	Coat condition	Ocular discharge
Back muscles	1	.388**	-.070	.023	.182*	.410**	-.278**	-.057	.235**	.181*	-.478**	-.006	.174*	.390**	-.051	.429**	-.136	.188*
Lameness	.388**	1	-.042	.181*	.023	.233**	-.229**	-.043	.289**	-.015	-.333**	-.082	.263**	.340**	.020	.288**	-.050	.235**
Low BCS	-.070	-.042	1	-.304**	.058	.069	-.125	.080	.079	.002	-.003	-.068	-.082	.049	.088	.056	.269**	-.003
High BCS	.023	.181*	-.304**	1	.103	.198*	-.042	-.172*	.023	.069	.006	.082	.104	-.036	-.031	.004	-.085	.127
HAT avoid	.182*	.023	.058	.103	1	.074	.037	-.052	.071	.098	-.171*	.070	.037	.111	-.021	.198*	.027	.048
Hoof condition	.410**	.233**	.069	.198*	.074	1	-.192*	.093	.264**	-.065	-.017	.033	.086	.102	.169*	.095	.246**	.304**
Skin irregularities	-.278**	-.229**	-.125	-.042	.037	-.192*	1	-.112	-.209*	-.047	.223**	.028	-.156	-.292**	-.134	-.225**	.136	-.068
NOT avoid	-.057	-.043	.080	-.172*	-.052	.093	-.112	1	.085	-.011	.121	.015	-.119	-.018	.204*	-.156	.032	-.113
Rubbed mane & tail	.235**	.289**	.079	.023	.071	.264**	-.209*	.085	1	-.110	.040	-.030	.229**	.216**	.198*	.102	.227**	.058
Stereotypies	.181*	-.015	.002	.069	.098	-.065	-.047	-.011	-.110	1	-.232**	.132	.010	.076	-.107	.280**	-.112	.025
White hairs	-.478**	-.333**	-.003	.006	-.171*	-.017	.223**	.121	.040	-.232**	1	.136	-.004	-.431**	.225**	-.542**	.115	-.066
Wounds (cuts)	-.006	-.082	-.068	.082	.070	.033	.028	.015	-.030	.132	.136	1	.028	.014	-.016	-.020	-.051	.125
Bars	.174*	.263**	-.082	.104	.037	.086	-.156	-.119	.229**	.010	-.004	.028	1	.122	.118	.229**	.006	.062
Mouth corners	.390**	.340**	.049	-.036	.111	.102	-.292**	-.018	.216**	.076	-.431**	.014	.122	1	-.006	.412**	-.139	.109
Nasal discharge	-.051	.020	.088	-.031	-.021	.169*	-.134	.204*	.198*	-.107	.225**	-.016	.118	-.006	1	-.032	.107	.039
Incisors	.429**	.288**	.056	.004	.198*	.095	-.225**	-.156	.102	.280**	-.542**	-.020	.229**	.412**	-.032	1	.043	.230**
Coat condition	-.136	-.050	.269**	-.085	.027	.246**	.136	.032	.227**	-.112	.115	-.051	.006	-.139	.107	.043	1	.046
Ocular discharge	.188*	.235**	-.003	.127	.048	.304**	-.068	-.113	.058	.025	-.066	.125	.062	.109	.039	.230**	.046	1

** Correlation is significant at the 0.01 level (2-tailed).* Correlation is significant at the 0.05 level (2-tailed).□

Validity of Animal Based measures

Table 4. References used to verify the validity of the animal based measures used in the Welfare Monitoring System Horses – assessment protocol version 2.0.

Animal based measures	References
Breathing frequency ¹	Evans et al., 1977; Geor et al., 1995; Hodgson and Rose, 1994; Murray et al., 2010; Pritchard et al., 2006; Pritchard et al., 2005
Breathing manner	Evans et al., 1977; Geor et al., 1995; Hodgson and Rose, 1994; Murray et al., 2010; Pritchard et al., 2006; Pritchard et al., 2005
White hairs	Murray et al., 2010; Sloet van Oldruitenborgh-Oosterbaan and Knottenbelt, 2001
Wounds	McGowan et al., 2010; Mejdell et al., 2010; Murray et al., 2010; Sloet van Oldruitenborgh-Oosterbaan and Knottenbelt, 2001
Generalized skin problem	Sloet van Oldruitenborgh-Oosterbaan and Knottenbelt, 2001
Rubbed and broken hairs mane and tail	Eriksson et al., 2008; Pilsworth and Knottenbelt, 2004; Sloet van Oldruitenborgh-Oosterbaan and Knottenbelt, 2001; van der Rijt et al., 2008
Itchiness	Fadok, 1995
Coat condition	Sloet van Oldruitenborgh-Oosterbaan and Knottenbelt, 2001
Body Condition Score	Carroll and Huntington, 1988; Christie et al., 2006; Coenen, 1998; Ellis, 2004; Johnson et al., 2009
Back muscles	Lesimple et al., 2010; Murray et al., 2010
Ear hairs ¹	Stull, 1999
Length whiskers	Ahl, 1986; Von Rotz and Friess, 1995
Ocular discharge	Barnett et al., 2004
Nasal discharge	McGorum and Dixon, 2007
Wear pattern incisors	Dixon, 2000; Van Lancker et al., 2007; Waran, 2002
Mouth corners	Baker and Easley, 2005; Tell et al., 2008; Van Lancker et al., 2007
Bars	Baker and Easley, 2005; Tell et al., 2008; Van Lancker et al., 2007
Hoof condition	Doughty et al., 2009; Florence and McDonnell, 2006; Kummer et al., 2009; Schüle and Appelbaum, 2003; Turner, 2003; Waran, 2002
Lameness	Broster et al., 2009; Burn et al., 2010a, b; Murray et al., 2010; Pritchard et al., 2005; Winkelsett and Vervuert, 2008
Coughing	McGorum and Dixon, 2007
Human Approach Test ¹	Hausberger and Muller, 2002; Keeling et al., 1999; Waiblinger et al., 2006; Waiblinger et al., 2002
Novel Object Test ¹	Visser et al., 2001; Wolff et al., 1997
Dry lying places	
Feeding lumps ¹	
Consistency of manure ¹	Burn et al., 2009; dos Santos et al., 2009; Smith et al., 1996; Zeyner et al., 2004

Full References

Ahl, A. S. 1986. The role of vibrissae in behavior: A status review. *Veterinary Research Communications* 10: 245-268.

Baker, G. J., and J. Easley. 2005. *Equine Dentistry*. 2 ed. Elsevier Saunders, London.

¹ These measures are not (yet) included in the full Welfare Quality model (version 2.0), but need to be registered at the farm when applicable.

- Barnett, K. C., S. M. Crispin, J. D. Lavach, and A. G. Matthews. 2004. *Equine Ophthalmology, an atlas & text*. Saunders, London.
- Broster, C. E., C. C. Burn, A. R. S. Barr, and H. R. Why. 2009. The range and prevalence of pathological abnormalities associated with lameness in working horses from developing countries. *Equine Vet J* 41: 474-481.
- Burn, C. C., T. L. Dennison, and H. R. Why. 2010a. Environmental and demographic risk factors for poor welfare in working horses, donkeys and mules in developing countries. *Vet. J.* 186: 385-392.
- Burn, C. C., T. L. Dennison, and H. R. Why. 2010b. Relationships between behaviour and health in working horses, donkeys, and mules in developing countries. *Appl. Anim. Behav. Sci.* 126: 109-118.
- Burn, C. C., J. C. Pritchard, and H. R. Why. 2009. Observer reliability for working equine welfare assessment: Problems with high prevalences of certain results. *Anim. Welf.* 18: 177-187.
- Carroll, C. L., and P. J. Huntington. 1988. Body condition scoring and weight estimation of horses. *Equine Vet J* 20: 41-45.
- Christie, J. L. et al. 2006. Management factors affecting stereotypes and body condition score in nonracing horses in Prince Edward Island. *Can. Vet. J.* 47: 136-143.
- Coenen, M. 1998. Body condition scoring of horses and cattle - A tool in veterinary services related to animal welfare. *Dtsch. Tierarztl. Wochenschr.* 105: 124-127.
- Dixon, P. M. 2000. Removal of equine dental overgrowths. *Equine Vet. Educ.* 12: 68-81.
- dos Santos, T. M. et al. 2009. Buffer capacity, pH and faeces consistency in horses submitted to dietetic starch overload. *Cienc. Rural* 39: 1782-1788.
- Doughty, A., N. Cross, A. Robins, and C. J. C. Phillips. 2009. The origin and foot condition of horses slaughtered in Australia for the human consumption market. *Equine Vet J* 41: 808-811.
- Ellis, A. D. 2004. *Paard en Voer - Praktische gids voor gezond voeren en beweiden*. Roodbont Uitgeverij, Zutphen.
- Eriksson, S. et al. 2008. Genetic analysis of insect bite hypersensitivity (summer eczema) in Icelandic horses. *Animal* 2: 360-365.
- Evans, J. W., A. Borton, H. F. Fintz, and L. D. VanVleck. 1977. *The Horse*. W.H. Freeman & Company, San Francisco.
- Fadok, V. A. 1995. Overview of equine papular and nodular dermatoses. *Vet Clin North Am Equine Pract* 11: 61-74.
- Florence, L., and S. M. McDonnell. 2006. Hoof growth and wear of semi-feral ponies during an annual summer 'self-trimming' period. *Equine Vet J* 38: 642-645.
- Geor, R. J., L. J. McCutcheon, G. L. Ecker, and M. I. Lindinger. 1995. Thermal and cardiorespiratory responses of horses to submaximal exercise under hot and humid conditions. *Equine veterinary journal. Supplement*: 125-132.
- Hausberger, A., and C. Muller. 2002. A brief note on some possible factors involved in the reactions of horses to humans. *Appl. Anim. Behav. Sci.* 76: 339-344.
- Hodgson, D. R., and R. J. Rose. 1994. *The athletic horse: principles and practice of equine sports medicine*. 1 ed. W. B. Saunders Company, Philadelphia.

- Johnson, P. J., C. E. Wiedmeyer, N. T. Messer, and V. K. Ganjam. 2009. Medical implications of obesity in horses - lessons for human obesity. *J Diabetes Sci Technol* 3: 163-174.
- Keeling, L. J., A. Blomberg, and J. Ladewig. 1999. Horse-riding accidents: when the human-animal relationship goes wrong! In: 33rd International Congress of the International Society for Applied Ethology, Lillehammer. p 86.
- Kummer, M., D. Gygax, C. Lischer, and J. Auer. 2009. Comparison of the trimming procedure of six different farriers by quantitative evaluation of hoof radiographs. *Vet. J.* 179: 401-406.
- Lesimple, C., C. Fureix, H. Menguy, and M. Hausberger. 2010. Human direct actions may alter animal welfare, a study on horses (*Equus caballus*). *PLoS ONE* 5: e10257.
- McGorum, B. C., and P. M. Dixon. 2007. Clinical examination of the respiratory tract. In: B. C. McGorum, P. M. Dixon, N. E. Robinson and J. Schumacher (eds.) *Equine respiratory medicine and surgery*. Saunders Elsevier, London.
- McGowan, T. W. et al. 2010. A survey of aged horses in Queensland, Australia. Part 2: Clinical signs and owners' perceptions of health and welfare. *Australian Veterinary Journal* 88: 465-471.
- Mejdell, C. M. et al. 2010. Reliability of an injury scoring system for horses. *Acta Vet. Scand.* 52.
- Murray, R. C., J. M. Walters, H. Snart, S. J. Dyson, and T. D. H. Parkin. 2010. Identification of risk factors for lameness in dressage horses. *Vet. J.* 184: 27-36.
- Pilsworth, R. C., and D. C. Knottenbelt. 2004. Equine insect hypersensitivity. *Equine Vet. Educ.* 16: 324-325.
- Pritchard, J. C., A. R. S. Barr, and H. R. Whay. 2006. Validity of a behavioural measure of heat stress and a skin tent test for dehydration in working horses and donkeys *Handbook of Environmental Chemistry, Volume 5: Water Pollution No. 38.* p 433-438.
- Pritchard, J. C., A. C. Lindberg, D. C. J. Main, and H. R. Whay. 2005. Assessment of the welfare of working horses, mules and donkeys, using health and behaviour parameters. *Preventive Veterinary Medicine* 69: 265-283.
- Schüle, E., and F. Appelbaum. 2003. The foot - Evaluation and interpretation of the findings during purchase examination. *Prakt. Tierarzt* 84: 682-689.
- Sloet van Oldruitenborgh-Oosterbaan, M. M., and D. C. Knottenbelt. 2001. *The practitioners guide to dermatology* Libre BV, Leeuwarden.
- Smith, B. L. et al. 1996. Effects of road transport on indices of stress in horses. *Equine Vet J* 28: 446-454.
- Stull, C. L. 1999. Responses of horses to trailer design, duration, and floor area during commercial transportation to slaughter. *J. Anim Sci.* 77: 2925-2933.
- Tell, A., A. Egenvall, T. Lundström, and O. Wattle. 2008. The prevalence of oral ulceration in Swedish horses when ridden with bit and bridle and when unriden. *Vet. J.* 178: 405-410.
- Turner, T. A. 2003. Examination of the equine foot. *Vet. Clin. North Am. Equine Pract.* 19: 309-332.
- van der Rijt, R., R. van den Boom, Y. Jongema, and M. M. S. van Oldruitenborgh-Oosterbaan. 2008. *Culicoides* species attracted to horses with and without insect hypersensitivity. *Vet. J.* 178: 91-97.

- Van Lancker, S., W. Van Den Broeck, and P. Simoens. 2007. Incidence and morphology of bone irregularities of the equine interdental spaces (bars of the mouth). *Equine Vet. Educ.* 19: 103-106.
- Visser, E. K. et al. 2001. Quantifying aspects of young horses' temperament: consistency of behavioural variables. *Appl. Anim. Behav. Sci.* 74: 241-258.
- Von Rotz, A., and A. E. Friess. 1995. A scanning electron-microscopic analysis of the morphology of equine lower lip sinus hair. *ACTA ANAT.* 154: 196-204.
- Waiblinger, S. et al. 2006. Assessing the human-animal relationship in farmed species: A critical review. *Appl. Anim. Behav. Sci.* 101: 185-242.
- Waiblinger, S., C. Menke, and G. Coleman. 2002. The relationship between attitudes, personal characteristics and behaviour of stockpeople and subsequent behaviour and production of dairy cows. *Appl. Anim. Behav. Sci.* 79: 195-219.
- Waran, N. 2002. *The welfare of horses.* Kluwer Academic Publishers, Dordrecht.
- Winkelsett, S., and I. Vervuert. 2008. Animal welfare in prophylaxis and therapy of laminitis. *Dtsch. Tierarztl. Wochenschr.* 115: 106-113.
- Wolff, A., M. Hausberger, and N. Le Scolan. 1997. Experimental tests to assess emotionality in horses. *Behavioural Processes* 40: 209-221.
- Zeyner, A., C. Geißler, and A. Dittrich. 2004. Effects of hay intake and feeding sequence on variables in faeces and faecal water (dry matter, pH value, organic acids, ammonia, buffering capacity) of horses. *J. Anim. Physiol. Anim. Nutr.* 88: 7-19.

Intra-observer reliability

Reliability between two trainers (veterinarians) for their assessment scores for 54 horses has been assessed on the animal based measures of the Welfare Monitoring System assessment protocol for horses (version 2.0). This study has been executed in July 2011.

Conclusion is based on the % of agreement between trainers for 108 observations (each horse is assessed twice) and the Cohen's kappa parameters. A kappa value of 0.00-0.20 is regarded as 'slight'(-), 0.21-0.40 as 'average' (\pm), 0.41-0.60 as 'moderate' (+), 0.61-0.80 as 'substantial' (++) and 0.81-1.0 as 'excellent' (+++) (Burn et al., 2009). For a number of measures the population of 54 horses was regarded as too homogenous for conclusive reliability ratings, these are depicted with a '?'.

Table 5. intra-observer reliability for animal based measures of the Welfare Monitoring System assessment protocol for horses (version 2.0).

Animal based measures	% agreement	Kappa value	Probability	Conclusion
Breathing frequency ²	75	N/A		++
Breathing manner	97	-0.014	0.510	?
White hairs				
Location 1	96	0.644	0.011	?
Location 2	97	0.555	0.067	?
Location 3	87	0.733	<0.001	++
Location 4	98	0.954	<0.001	++
Location 5	98	-0.011	0.506	?
Wounds				
Location 1	89	0.408	0.021	+
Location 2	78	0.299	0.017	\pm
Location 3	96	0.317	0.215	?
Location 4	89	0.398	0.024	\pm
Location 5	77	0.451	<0.001	+
Location 6	89	0.600	<0.001	+
Location 7	91	0.716	<0.001	++
Location 8	82	0.593	<0.001	+
Location 9	100	1.000	0.022	?
Generalized skin problem	98	-0.009	0.505	?
Rubbed and broken hairs mane and tail	90	0.541	0.001	+
Itchiness				?
Coat condition	99	-0.005	0.502	?
Body Condition Score	70	0.485	<0.001	+
Back muscles	78	0.442	<0.001	+
Ear hairs ²	93	0.294	0.148	\pm
Length whiskers	94	-0.029	0.531	?
Ocular discharge	83	0.657	<0.001	++
Nasal discharge	99	0.662	0.124	?
Wear pattern incisors				
Feed intake	82	0.644	<0.001	++
Crib-biting	94	0.221	0.262	?
Mouth corners	86	0.685	<0.001	++
Bars	90	0.037	0.447	?
Hoof condition	99	-0.005	0.502	?
Lameness	76	0.400	<0.001	+
Coughing	100	1.000	0.002	?

² These measures are not (yet) included in the full Welfare Quality model (version 2.0), but need to be registered at the farm when applicable.

Animal based measures	% agreement	Kappa value	Probability	Conclusion
Human Approach Test ²	87	0.729	<0.001	++
Novel Object Test ²	83	0.738	<0.001	++
Dry lying places	99	0.622	0.123	?
Feeding lumps ²				?
Consistency of manure ²	85	0.760	<0.001	++

Inter-observer reliability

Reliability within two trainers (veterinarians) for their assessment scores for 54 horses. To test for the within reliability all 54 horses have been assessed twice. For the assessment of the animal based measures the Welfare Monitoring System assessment protocol for horses (version 2.0) has been used. This study has been executed in July 2011.

Conclusion is based on the % of agreement within a trainer (for one measure) and the Cohen's kappa parameters for the other measures. A kappa value of 0.00-0.20 is regarded as 'slight'(-), 0.21-0.40 as 'average' (\pm), 0.41-0.60 as 'moderate' (+), 0.61-0.80 as 'substantial' (++) and 0.81-1.0 as 'excellent' (+++) (Burn et al., 2009). For a number of measures the population of 54 horses was regarded as too homogenous for conclusive reliability ratings, these are depicted with a '?'.

Table 6. inter-observer reliability for animal based measures of the Welfare Monitoring System assessment protocol for horses (version 2.0).

Animal based measures	% agreement	Kappa value	Probability	Conclusion
Breathing frequency ³	87	N/A		++
Breathing manner	N/A	-0.014	0.510	?
White hairs	N/A			
Location 1		0.644	0.011	?
Location 2		0.555	0.067	?
Location 3		0.777	<0.001	++
Location 4		0.954	<0.001	+++
Location 5		-0.011	0.506	?
Wounds	N/A			
Location 1		0.556	0.003	+
Location 2		0.737	<0.001	++
Location 3		-0.025	0.524	?
Location 4		0.649	<0.001	++
Location 5		0.735	<0.001	++
Location 6		0.800	<0.001	++
Location 7		0.687	<0.001	++
Location 8		0.756	<0.001	++
Location 9		1.000	0.022	?
Generalized skin problem	N/A	1.000	0.078	?
Rubbed and broken hairs mane and tail	N/A	0.750	<0.001	++
Itchiness				?
Coat condition	N/A	-0.005	0.502	?
Body Condition Score	N/A	0.642	<0.001	++
Back muscles	N/A	0.368	0.007	\pm
Ear hairs ³	N/A	0.471	0.047	+
Length whiskers	N/A	0.599	0.066	?
Ocular discharge	N/A	0.162	0.050	?
Nasal discharge	N/A	-0.014	0.510	?
Wear pattern incisors	N/A			
Feed intake		0.682	<0.001	++
Crib-biting		-0.380	0.544	?
Mouth corners	N/A	0.601	<0.001	++
Bars	N/A	0.650	0.010	?
Hoof condition	N/A	-0.005	0.502	?
Lameness	N/A	0.492	<0.001	+
Coughing	N/A	0.481	0.083	?

³ These measures are not (yet) included in the full Welfare Quality model (version 2.0), but need to be registered at the farm when applicable.

Animal based measures	% agreement	Kappa value	Probability	Conclusion
Human Approach Test ³	N/A	-0.038	0.638	?
Novel Object Test ³	N/A	0.243	<0.001	±
Dry lying places	N/A	-0.130	0.509	?
Feeding lumps ³				?
Consistency of manure ³	N/A	0.800	<0.001	++

