Archived at http://orgprints.org/15915

Results of the CoreOrganic-Workshop on animal based parameters in Trenthorst, Germany (04.02.08-08.02.2008) *Solveig March, Lisi Gratzer, Jan Brinkmann, Christoph Winckler*

Objectives and methods

It was the aim of this meeting to train the assessors of several CoreOrganic-Partners (i.e. Austria, Germany, Denmark and Norway) with regard to the methods which are going to be applied in the course of the project on-farm assessments. Training involved animal-based parameters, a resource checklist and a management questionnaire.

After a theoretical and practical introduction, inter-observer reliability (IOR) of the main animal-based parameters was tested. This included gait scoring/lameness, body condition, cleanliness and integument alterations. The scoring systems for gait scoring, cleanliness and integument alterations had been adapted from the WelfareQuality® protocol; body condition was scored according to Metzner et al. (1993).

Subjective scoring systems which are used in many epidemiological studies have the advantage that they do not require any equipment but - due to the subjectivity of the assessment - it is necessary to achieve acceptable inter-observer reliability (IOR) before and after the study to ensure valid data.

To estimate the inter-observer reliability, the prevalence-adjusted bias-adjusted Kappa (PABAK) as well as the proportion of agreements between the "gold standard" of an experienced assessor and all untrained assessors were calculated. The PABAK(= [(k*p)-1]/(k-1) where k=number of categories and p=proportion of matchings) is based on the unweighted Cohens kappa test. According to Byrt et al. (1993), the Kappa coefficient measures the agreement beyond what would be expected by chance. The weighted Kappa coefficient also takes into account, that larger disagreement is more important than near disagreement. Finally, the prevalence-adjusted bias-adjusted Kappa (PABAK) is the value that kappa would take if, in addition, the prevalence of each category was equal (Gunnarson et al. 2000).

All coefficients may range between 0 and 1 meaning no agreement between observations if the coefficient is equal to 0 and perfect agreement if the value is equal to 1. Matchings are only counted, if both observers give exactly the same score and PABAK can reach values 0 to 1: values above 0 show a positive correlation between observer's ratings.

Calculations were carried out for the following scoring-systems (see on-farm assessment protocol Aniplan):

- Locomotion-scoring (3-score-system: 0 – Not Lame: Timing of steps and weight-bearing equal on all four feet; 1 – Lame: Imperfect temporal rhythm in stride creating a limp; 2 – Severely Lame: Strong reluctance to bear weight on one limb, or more than one limb affected),
- Body condition scoring / BCS (original 17-score key (min=1.00, max=5.00, 0.25 intervals / modified 5-score-system if accepting a deviation of +/- 0.25),
- Cleanliness scoring of four body regions (lower hind legs, hind quarter and flank, udder and teats): 2-category-scoring-system (acceptable/dirty) and for the teats: 3-category-scoring system (acceptable/any dirt present/large plaques of dirt).Integument alterations of three different categories were counted in four body regions (carpus, tarsus, neck, flank; "hairless patch": area with hair loss or extensively thinned hair as a response to parasites, skin not damaged, hyperkeratosis possible; "lesion": area with damaged skin either in form of a scab or a wound, dermatitis due to ectoparasites or (partly) missing teats; "swelling": overt swellings). For the calculation of IOR a binomial score was used (described alteration present in each location or not present).
- IOR testing was carried out on-farm, except for locomotion scoring, which was also done with video material.

With regard to the acceptability of the level of agreement, Fleiss et al. (2003) described PABAK values > 0.6 – 0.8 as an expression of a good/satisfactory agreement, and values > 0.8 as very satisfactory agreement. Holzhauer et al. (2004) defined Kappa values between 0.4 and 0.5 as moderate, values between 0.5 and 0.6 as sufficient and values between 0.6 and 0.8 as good. Accordingly, PABAK values lower than 0.4 are rated as unsatisfactory; values above 0.4 as acceptable, above 0.6 as good/satisfactory and above 0.8 as very good (Keppler et al. 2004).

Results

Locomotion scoring

For IOR testing of gait scoring in total 53 cows were assessed, out of which 20 "live" on-farm and 33 using video clips. The range of proportion of agreement for all cows was 0.74 to 0.94, and PABAK values ranged between 0.60 and 0.92. Scores from video clips generally showed lower inter-observer agreement (Table 1).

Table 1: Inter-observer reliability of locomotion scoring

	PABAK	proportion of agreement
Total, n=53 (including 33 video clips)	0.71 (0.60 – 0.92)	0.81 (0.74 – 0.94)
without video clips (n=20)	*0.84* *(0.70 – 0.93)*	*0.89* *(0.80 – 0.95)*

Table 2: Inter-observer reliability of body condition scoring

n= 20	PABAK (mean / min – max)
BCS (5 categories: Δ +/- 0.25 tolerance)	0.79 0.69 – 0.94
BCS (17 categories: exact agreement)	*0.35* *0.20 – 0.52*

The results of the inter-observer reliability for scoring the other animal based parameters (BCS, integument alterations and cleanliness) are shown in tables 2-4.

Table 3: Inter-observer reliability for scoring integument alterations

	PABAK (mean / min – max)		
n=20	Hairless patches	Lesions	Swellings
Carpus	0.61 0.40 – 0.90	0.69 0.50 – 0.80	0.75 0.50 – 0.90
Tarsus	0.39 0.10 – 0.50	0.73 0.60 – 0.90	0.84 0.70 – 1.00
Neck	0.64 0.40 – 0.90	*Marginal prevalence*	0.90 0.80 – 1.00
Flank	0.79 0.70 – 0.90	0.86 0.70 – 1.00	*Marginal prevalence*

Table 4: Inter-observer reliability of cleanliness scoring

n= 20	PABAK (mean / min – max)
Legs	0.59 0.30 – 0.90
Flank	0.75 0.40 – 1.00
Udder	0.85 0.50 – 1.00
Teats	0.65 0.48 – 0.85

Conclusions

Taking the PABAKs between untrained and experienced assessors (set as "gold standard") into account, at least satisfactory levels of inter-observer reliability were found for all parameters. This was achieved after theoretical and practical introduction into the scoring systems. The observer training will therefore allow for comparable and reliable data sets gathered in all European CoreOrganic subprojects.

References

Fleiss JL, Levin B, Paik MC (2003): Statistical Methods for Rates and Proportions. John Wiley & Sons, New York, USA, in Woodward, M., 2005: Epidemiology - Study Design and Data Analysis, chapter 2: Basic Analytical Procedures, p.99. 2nd ed. Chapman & Hall/CRC, Boca Raton, USA.

Gunnarson S (2000): Laying hens in loose housing systems. Doctoral thesis, Swedish University of Agricultural Sciences: Uppsala, Sweden.

Holzhauer M, Middelesch H, Bartels C, Frankena K (2004): Evaluation of a dutch claw health scoring system in dairy cattle. Proceedings of the 13th international sysmposium and 5th conference on lameness in ruminants: 18-21.

Keppler C, Schubbert A, Knierim U (2004): Welche Methoden sind zur Beurteilung von Hühnern im Hinblick auf Federpicken und Kannibalismus geeignet? Erste Untersuchungen zum Vergleich verschiedener Methoden im Hinblick auf Durchführbarkeit, Aussagekraft und Wiederholbarkeit. 11. Freilandtagung und 17. IGN-Tagung: 71-74.

Metzner M, Heuwieser W, Klee W (1993): Die Beurteilung der Körperkondition (body condition scoring) im Herdenmanagement. Prakt Tierarzt 11: 991 – 998.