INFLUENCE OF SAMPLE SIZE AND EXPERIMENTER ON RELIABILITY OF MEASURES OF AVOIDANCE DISTANCE IN DAIRY COWS

S Waiblinger* and C Menke

Institute of Animal Husbandry and Animal Welfare, University of Veterinary Medicine, Veterinärplatz 1, A-1210 Vienna, Austria

* Contact for correspondence and requests for reprints: Susanne.Waiblinger@vu-wien.ac.at

Abstract *Animal Welfare 2003, **12**: 585-589*

In the present study we evaluated the influence of different sample sizes and different experimenters on the reliability of measures of avoidance distance (AD) at farm level. On 29 dairy farms the AD of 55–100% of the cows was assessed by two different experimenters (E1 and E2). For both experimenters the herd median of AD (ADME) and the percentage of animals that could be touched (Touch%) were calculated. The reliability between experimenters was assessed by Spearman rank correlation coefficients. To assess the influence of sample size on reliability of AD, the tested animals were randomly divided into two halves (H1 and H2), and ADME and Touch% were calculated for both halves and correlated with each other, with total ADME and total Touch%, and with the behaviour of the milkers. All measures of AD were highly correlated between experimenters (ADME $r_s = 0.86$; Touch% $r_s = 0.81$). On farms with a higher value for ADME, however, some discrepancy was found between experimenters in ADME and Touch%. Smaller sample size reduced the number of significant correlations with milkers' behaviour. AD of H1 and H2 correlated only moderately ($r_s = 0.38–0.43$). In sum, smaller sample size reduced reliability and validity. Between-observer reliability of AD was relatively high, but there may be some observer influence. Further investigations are necessary to optimise the measures.

Keywords: *animal welfare, avoidance distance, cattle, human–animal relationship, on-farm welfare assessment*

Introduction

The human–animal relationship can have a major impact on animal health, production and welfare (Menke *et al* 1999; for a review, see Hemsworth & Coleman 1998). Therefore, a complete welfare assessment on farms should include this factor. Reactions of animals toward humans can be used to directly assess the human–animal relationship from the animals' perspective, reflecting the animals' level of fear of or confidence in humans. Avoidance distance (AD) of cows toward humans in the barn has been shown to correlate well with stockperson behaviour, supporting the validity of this measure (Waiblinger *et al* 2002). However, little is known about the reliability of measures of AD, which is influenced by differences between observers in the exact way of testing and in assessment of distance (inter-observer reliability), consistency of the animals in their AD toward the same human (test–retest reliability) and consistency of avoidance reactions toward different humans.

Compared with other reactions to humans, such as approach behaviour in a test pen, AD is relatively easy to assess in dairy herds but, depending on herd size, it can take a considerable

length of time. Thus, sample size affects the feasibility of the test. Reducing the sample size saves time but may decrease the reliability of the measure and, if the measure taken is not a true reflection of the herd value, this also influences validity. In the present study we evaluated the influence of different sample sizes and of different experimenters on the reliability of measures of AD at farm level.

Methods

On 29 dairy farms (22–50 cows per herd; loose housing with cubicles), the AD of 55–100% of the cows in the barn was assessed by two different experimenters, E1 and E2, both dressed in green overalls (E1 female, E2 male). Standing animals were approached slowly (one step per second) from the front, and the distance between the person's hand and the animal's head was estimated at the moment of withdrawal (stepping away or turning away of the head). Because it was intended that the test also be used to compare cows' reactions toward an unknown person between farms, the procedure needed to be equal on all farms for experimenter 1. The procedure was as follows: E1 always began by assessing AD in all standing cows; when this was complete, E2 began the same procedure. The intention was to test at least 70% of the animals. On 12 farms this was achieved in a single session. On 17 farms, two measurement sessions were necessary. For both E1 and E2, the herd median for AD (ADME, measured in m) and the proportion of animals that could be touched (Touch%; in % of tested animals) were calculated for all tested cows, and labelled ADME-all and Touch%-all. This procedure inevitably meant that E1 and E2 did not test exactly the same cows or the same proportion of the herd. Thus, to exclude any additional influences caused by testing different cows or by different sample sizes, we also calculated the measures for only those cows that were tested by both experimenters; these were labelled ADME-same and Touch%-same. The proportion of cows tested by E1 and also by E2 varied on the farms from 47% to 82% of all cows in the herd. The reliability between experimenters was assessed by Spearman rank correlation coefficients.

To assess the influence of sample size on reliability of AD, the animals tested by E1 (70–100% of the cows) were randomly divided into two halves (H1 and H2). ADME and Touch% were calculated for both halves and correlated (Spearman) with each other (ie split-half correlation) and with total ADME and total Touch% (all tested cows included). Furthermore, to investigate the influence of reduced sample size on validity, correlations with the behaviour of the milkers were calculated. The milkers' behaviour was observed during one evening milking. Interactions with cows were grouped together into positive (petting, touching, talking quietly), neutral (gentle hitting with hand/stick, talking dominantly) and negative (forceful hitting with stick/hand, talking impatiently, shouting), and the total number per cow (Pos, Neu, Neg) as well as the percentage of all interactions (Pos%, Neu%, Neg%) were calculated.

Results

Differences between experimenters

Measures of AD were highly correlated between experimenters (all animals tested: ADME-all $r_s = 0.86$, Touch%-all $r_s = 0.81$; animals that were tested by both experimenters: ADME-same $r_s = 0.85$, Touch%-same $r_s = 0.75$; each $P = 0.000$; n = 29; see Figure 1). Discrepancy in ADME between experimenters was particularly found on farms with a higher value for ADME. There was a directed effect in Touch%: on most farms, Touch% was higher for E2 than for E1 (20 farms), on six farms there was no difference and on three farms Touch% was

lower for E2. On some farms, the difference between Touch% of the two experimenters was high (Touch%-sameE1 minus Touch%-sameE2: median –11%; range +6% to –32%).

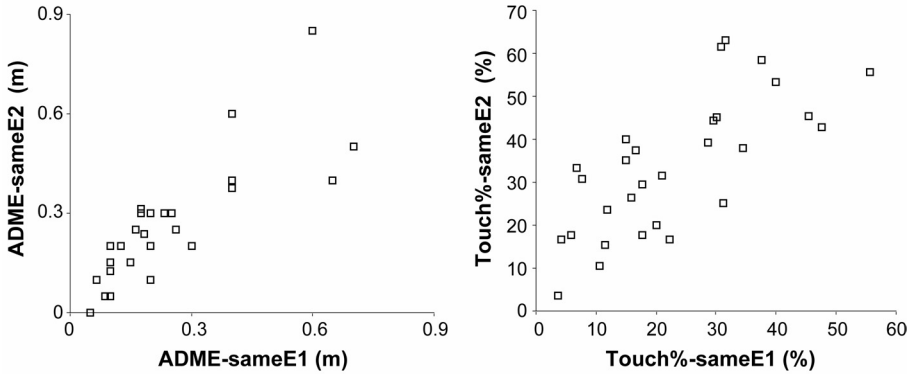

Figure 1 **Scatter plots of measures of avoidance distance (ADME and Touch%) of the two experimenters E1 and E2 in the animals tested by both.**

Sample size

Because of their contribution to the total value, correlations of the two halves H1 and H2 with the total were relatively high (ADME: H1-total $r_s = 0.76$, H2-total $r_s = 0.80$; Touch%: H1-total $r_s = 0.88$, H2-total $r_s = 0.76$; $P = 0.000$ for each), but the split-half correlations were only moderate (ADME H1-H2: $r_s = 0.38$, $P < 0.05$; Touch% H1-H2: $r_s = 0.43$, $P < 0.05$). Accordingly, reducing the sample size (within one experimenter) to half of the cows lowered correlation coefficients (except in two out of 24 coefficients) and reduced the number of significant correlations with milkers' behaviour (Table 1). The effects were different for the two halves and also were not consistent for ADME and Touch%, but the differences were not statistically significant.

Table 1 **Spearman correlations between the behaviour of milkers and measures of avoidance distance of all tested cows (total) or of half of the cows (H1, H2). n = 29 farms. ADME, herd median of avoidance distance; Touch%, percentage of animals that could be touched.**

		ADME			Touch%		
		H1	**H2**	**total**	**H1**	**H2**	**total**
Behaviour of milkers	Pos	–0.24	–0.46*	–0.53**	0.40*	0.26	0.44*
	Pos%	–0.30	–0.40*	–0.54**	0.44**	0.28	0.47*
	Neu	0.05	0.03	0.15	–0.13	–0.24	–0.18
	Neu%	0.16	0.27	0.42*	–0.25	–0.34	*–0.36*
	Neg	0.40*	0.38*	0.47**	*–0.35*	–0.23	–0.36
	Neg%	0.39*	0.39*	0.46**	*–0.31*	–0.13	*–0.31*

* $P < 05$ (2-tailed)
** $P < 01$ (2-tailed)
In italics, $P < 0.1$ (2-tailed)

Discussion

Differences between experimenters

The reactions of cows toward humans reflect their previous experience with humans. Cattle generalise their experiences with one human to others (Boivin *et al* 1992; Hemsworth *et al*

https://doi.org/10.1017/S0962728600026221 Published online by Cambridge University Press

1989, 1996). This can be used in on-farm assessment of the animals' relationship with humans by testing the reactions to an unknown person. Here, a generalised response to unknown people is expected. Our results are in line with this expectation and previous studies. However, although similar, there were some differences in AD between experimenters. Although they generalise, cattle are also able to discriminate between people (Boivin *et al* 1998; Rybarczyk *et al* 2001) and they differ in their reactions according to previous handling (Munksgaard *et al* 1997). Several factors could have contributed to the differences in avoidance reactions to the two experimenters. Experimenter 1 always began the testing of the cows. Cows probably become used to strangers after a while and subsequently their AD diminishes slightly. A greater degree of habituation might be the underlying basis of the lower AD measured by experimenter 2. However, previous results on consistency of cows' avoidance reactions toward one unknown person (Rousing & Waiblinger 2002) suggest at most a marginal effect of the order of testing. Subtle differences between the experimenters in the exact performance of the AD test (different gaze direction, speed of approach, or posture) might affect the cows' reactions. Also, physical characteristics of experimenters probably influence the reactions. The farms were family-run, and in most cases both the farmer and his wife worked with the cows, the women most often being responsible for milking and the men for feeding (and partly milking). It could be that cows had different experiences with men and women and generalised these to strangers of the respective gender. Cows use body shape to differentiate between people (Rybarczyk *et al* 2001) and dissimilarities between men and women are ubiquitous. Further studies are needed to investigate in further detail reactions toward people differing in various characteristics.

Sample size

The smaller the sample size required for a reliable and valid assessment of AD, the more feasible the assessment is in terms of reduced length of time needed. However, in the present study, reducing the sample size to less than 50% did lead to a substantial decrease in reliability and validity compared to a sample size of 70–100% of cows. Interestingly, the correlations between ADME and negative stockperson behaviour remained significant. Further investigations and discussions are required to determine the boundaries of compromise between feasibility and satisfactorily reliable assessment. Furthermore, sample size and sample selection must also be investigated in large dairy farms.

Conclusions and animal welfare implications

To improve animal welfare on farms it is necessary to reliably reveal the problematic aspects. Among other things, assessment of the human–animal relationship is required. Smaller sample sizes reduce the reliability and validity of this assessment. Between-observer reliability of AD was relatively high, but there may be some observer influence on the reactions of cows. Further studies are necessary to reveal the causes of this difference and the necessary sample size, and to optimise methods with regard to feasibility, reliability and validity.

Acknowledgements

We would like to thank all the farmers for their participation in this study.

References

Boivin X, Garel J P, Mante A and LeNeindre P 1998 Beef calves react differently to different handlers according to the test situation and their previous interactions with their caretaker. *Applied Animal Behaviour Science 55*: 245-257

Boivin X, LeNeindre P, Chupin J M, Garel J P and Trillat C 1992 Influence of breed and early management on ease of handling and open-field behaviour of cattle. *Applied Animal Behaviour Science 32*: 313-323

Hemsworth P H and Coleman G J 1998 *Human–Livestock Interactions: The Stockperson and the Productivity and Welfare of Intensively Farmed Animals*. CAB International: New York, USA

Hemsworth P H, Barnett J L, Tilbrook A J and Hansen C 1989 The effects of handling by humans at calving and during milking on the behaviour and milk cortisol concentrations of primiparous dairy cows. *Applied Animal Behaviour Science 22*: 313-326

Hemsworth P H, Price E O and Borgwardt R 1996 Behavioural responses of domestic pigs and cattle to human kind novel stimuli. *Applied Animal Behaviour Science 50*: 43-56

Menke C, Waiblinger S, Fölsch D W and Wiepkema P R 1999 Social behaviour and injuries of horned cows in loose housing systems. *Animal Welfare 8*: 243-258

Munksgaard L, DePassillé A M, Rushen J, Thodberg K and Jensen M B 1997 Discrimination of people by dairy cows based on handling. *Journal of Dairy Science 80*: 1106-1112

Rousing T and Waiblinger S 2002 Evaluation of on-farm methods for testing the human–animal relationship in dairy herds with cubicle loose housing systems — test–retest and inter-observer reliability and consistency to familiarity of test person. In: Rousing T 2002 *Welfare assessment in dairy herds with loose-housing cubicle systems — development and evaluation of welfare indicators* pp 30-52. PhD thesis, The Royal Veterinary and Agricultural University, Copenhagen, and Research Centre Foulum, Denmark

Rybarczyk P, Koba Y, Rushen J, Tanida H and De Passillé A M 2001 Can cows discriminate people by their faces? *Applied Animal Behaviour Science 74*: 175-189

Waiblinger S, Menke C and Coleman G 2002 The relationship between attitudes, personal characteristics and behaviour of stockpeople and subsequent behaviour and production of dairy cows. *Applied Animal Behaviour Science 79*: 195-219