



## Ranking experts' preferences regarding measures and methods of assessment of welfare in dairy herds using Adaptive Conjoint Analysis

J. J. Lievaart<sup>\*1,2</sup> and J. P. T. M. Noordhuizen<sup>†‡</sup>

<sup>\*</sup>University of Utrecht, Department of Farm Animal Health, Faculty of Veterinary Medicine, Utrecht, the Netherlands

<sup>†</sup>VACQA-International Consultancy, Santarém, Portugal

<sup>‡</sup>Population Medicine Group of the National Veterinary School of Lyon, Vet-Agro Sup campus, Lyon, France

### ABSTRACT

Welfare in dairy herds can be addressed using different concepts. The difficulty is to extract which measures are the most important to practically address welfare at the herd level and the methods to assess traits considered most important. Therefore, the preferences of 24 acknowledged European welfare experts were ranked regarding 70 measures suitable to assess dairy cattle welfare at herd level using the Adaptive Conjoint Analysis (ACA; Sawtooth Software, Inc., Sequim, WA) technique. The experts were selected on the basis of 3 criteria: at least 5 yr experience in animal welfare research; recent scientific publications in the field of animal welfare; and, at the most, 3 animal species including dairy cattle as their field of expertise. The 70 traits were ranked by using the median ACA questionnaire utility scores and the range between the answers of the 24 experts. A high utility score with a low range between the answers of the experts was considered as suitable to assess welfare at farm level. Measures meeting these criteria were prevalence of lameness cases ( $107.3 \pm 11.7$ ), competition for feed and water ( $96.4 \pm 13.9$ ), and number of freestalls per 10 cows ( $84.8 \pm 13.3$ ). Based on the utility score alone, these former measures were replaced by stereotypic behavior ( $111.7 \pm 17.1$ ), prevalence of lameness cases ( $107.3 \pm 11.7$ ), body condition score ( $108.0 \pm 18.9$ ), and hock lesions ( $104.7 \pm 16.1$ ). Subsequently, to demonstrate that the ACA technique can be used to rank either well-known or inconclusive methods of assessment, the methods for the traits lameness cases and the hygiene of the calving pen were ranked using another 2 ACA questionnaires. The results are based on the opinions of selected, internationally acknowledged dairy cattle welfare experts

within the European Union. In the future, other parties like dairy farmers and farmers' organization should be included to achieve consensus about the most suitable traits applicable in practice. The currently investigated traits do not always apply to all dairy husbandry systems across the world, but are based on a system that includes indoor housing during winter. It is concluded that ACA is a useful technique to rank the different scientific opinions of experts regarding suitable traits and methods of assessment of dairy cattle at the herd level.

**Key words:** dairy herd, welfare, assessment, adaptive conjoint analysis

### INTRODUCTION

Increasing consumer concern about farm animal welfare and legislation regarding animal welfare emphasize the need to develop a practical welfare assessment protocol that is feasible for the industry itself and sufficiently transparent for consumers (Winckler et al., 2003; Frewer et al., 2005; Webster, 2005). Over the last 2 decades, several concepts have been developed such as the concept of the Five Freedoms in the UK (FAWC, 1992; Webster, 2001), the Animal Needs Index (ANI) or Tiergerechtigkeitsindex (TGI) in Austria (Bartussek, 1999), and the concept of Biological Needs (Bracke et al., 2002). Measures used in the various concepts vary from possibility of mobility, social contact, flooring, climatization and human care, to the observations of animals (Capdeville and Veissier, 2001; Main et al., 2001; Main et al., 2003), and direct animal-based measurements and farm records analysis (Whay et al., 2003b). As a result, it is difficult to find consensus among welfare experts on the most important measures and their methods of assessment.

The Adaptive Conjoint Analysis (ACA; Sawtooth Software, Inc., 2002; Sequim, WA) technique was used for the ranking of welfare observations by interviewing several acknowledged European dairy welfare experts. The ACA technique converts the preferences of the experts into a quantitative utility score (US), and subsequently tests the consistency in the expert's answers.

Received October 21, 2010.

Accepted March 8, 2011.

<sup>1</sup>Corresponding author: jlievaart@csu.edu.au

<sup>2</sup>Current address: EH Graham Centre for Agricultural Innovation (an alliance between Charles Sturt University and NSW Department of Primary Industries), Locked Bag 588, Wagga Wagga NSW 2678, Australia.

The objectives were first to determine if the method of ACA is a useful method to rank expert preferences. Second, the suitability of the technique was examined to rank methods of assessment, which varied between methods that are well known and well described in the literature to ones that are less known and have no published references.

## MATERIALS AND METHODS

### *Study Design*

The study started with the selection of measures and welfare experts, as described in the section Selection of Measures, Welfare Experts, and Methods of Assessment. Subsequently, a first ACA questionnaire (ACA questionnaire 1) was composed in which all traits were to be ranked by animal welfare experts. The outcome of questionnaire 1 represented US for each trait, which were ranked and divided into 3 categories: low, medium, and high US. The design of the ACA questionnaire and the method of ranking of the measures are described in the section The ACA Design and Ranking of the Measures and Methods of Assessment. After ACA questionnaire 1, 2 subsequent questionnaires were sent on how to assess a particular individual observation at herd level to those experts who fitted the inclusion criteria of ACA questionnaire 1 and completed ACA questionnaire 1 with a sufficient correlation coefficient ( $>50\%$ ).

In the second and third questionnaire (ACA questionnaire 2 and 3) the method preferred by experts for measuring 2 traits (lameness cases and hygiene in the calving pen) were examined in more detail by, again, using the ACA technique. Both measures were selected from the category with relative high US ( $US \geq 70$ ), but variation exists in how well the methods of assessment are described in literature. The lameness cases trait has various well-established methods of assessment, which is in contrast to hygiene in the calving pen, which has different methods of assessment. The reason for choosing these 2 traits was to demonstrate that the ACA technique could be used to rank methods of assessment from either well-established or inconclusive methods of assessment. In ACA questionnaire 2, 9 methods for assessing lameness cases were evaluated, and in ACA questionnaire 3, 8 possible methods for the assessment of the hygiene in the calving pen were evaluated.

### *Selection of Measures, Welfare Experts, and Methods of Assessment*

Initially, all measures of which data are available at cow level, management (including procedures) and the

environment were considered included in all 3 questionnaires. Subsequently, the measures used in ACA questionnaire 1 and 2 had to meet the following 3 selection criteria: (1) dairy cattle orientated, (2) published in double-refereed journals, and (3) conformed to European legislation on the housing and management of dairy cattle (European Commission, 2006). This last criterion resulted in excluding measures, such as tail docking or the use of bovine somatotropin. For ACA questionnaire 3, only the first and the third selection criteria could be met.

To contact a maximum number of suitable welfare experts, the board of a European research consortium called Measuring and Monitoring Farm Animal Welfare (EU financial grants; Blokhuis et al., 2006) was contacted and provided a list of names and addresses of internationally acknowledged European welfare experts. In total, 79 experts were identified who, upon invitation and acceptance, received an access code to access the online ACA questionnaire 1, together with an introductory letter to explain the rationale behind this questionnaire. To accomplish a uniform profile of the experts with respect to the level of expertise, the questionnaire started with several open questions such as: years of experience, field of expertise, and list of publications. Experts were included if they met the following 3 inclusion criteria: 1) at least 5 yr of experience in the field of animal welfare; 2) expertise with a maximum of 3 species, including dairy cattle; 3) a minimum of 1 animal welfare-related publication as a first author in a peer-reviewed journal in the last 5 yr, or an animal welfare-related publication as a co-author, in a peer-reviewed journal in the last 5 yr.

In addition to these selection criteria, the consistency of the answers of the individual expert was determined. This was done using the same ACA software to produce a correlation coefficient ( $r$ ) for each expert. A threshold value of 50% was used for  $r$ , on the basis of the literature (van Schaik et al., 1998; Valeeva et al., 2005). This was underpinned by a sensitivity analysis of  $r$  by testing various threshold values between 0 and 60% with 5% intervals. The  $r$  of 50% or more of individual experts did not influence the overall ranking of measures.

### *The ACA Design and Ranking of the Measures and Methods of Assessment*

The ACA questionnaire examined personal preferences regarding certain components of a product, and created a quantitative US from the result. Each component has several choices, known as levels, within this program (ACA Systems Version 5.2; Sawtooth Software Inc., 2005). An example could be to test consumer pref-

erences when buying a new car. This product car has several components, such as color, which could contain 3 levels: red, yellow, and blue.

A respondent's utility is a measurement of his or her relative strength of preference for a given welfare measure or method of assessment. The absolute values of the utilities have no inherent meaning. The US itself reflects a score of one measure or method of assessment of animal welfare relative to the others and was used to rank the measures or method of assessment.

In ACA questionnaire 1, the product animal welfare was divided into 11 components (feed and water, excretion, resting and lying, thermoregulation, locomotion, body care, social comfort, reproduction, health, safety, and other traits) deduced from the Five Freedoms related to animal welfare, and the biological needs, which is another concept of animal welfare in sows (Bracke et al., 2002; Webster, 2003). All 70 measures selected were divided over these 11 components to represent the levels belonging to each component. Therefore, all measures were assessed by the experts.

In ACA questionnaire 2 on the assessment of lameness cases, the 9 methods were divided over 3 components: gait scoring methods, clinical findings, and record keeping. In ACA questionnaire 3, 8 methods were divided over 2 components: visual cleanliness and hygiene.

All ACA questionnaires were divided into 4 sections with different types of questions to refine the US per measure or method of assessment, and examine the consistency (called the correlation coefficient in the ACA program) between the answers of each respondent. The ACA program rescales the US from each trait of the questionnaire to a point scale. For further details on the adaptive conjoint technique and the  $r$ , refer to the ACA User Manual Version 5.2 (Sawtooth Software Inc., 2005). Each questionnaire ended with an open question asking whether any measures were missing.

As a consequence of the non-normal distribution of the outcome, the median instead of the mean per individual trait and per method of assessment are presented. To examine the degree of agreement between the experts, the absolute range per trait and per method of the US was calculated. To combine the US and the range per measure or per method of assessment, all US were first divided into a low ( $\leq 35$ ), median (35 to 70), or high ( $\geq 70$ ) US category, respectively, to create 3 categories with almost equal numbers of traits. Subsequently, for each category (low, medium, and high utility), the traits were again divided into 3 groups based on their absolute range between the answers of the experts. This time, the values of  $\leq 15$ , 15 to 30, and  $\geq 30$  were set as the threshold values to indicate a small, medium, or large range between the answers of the experts, respectively. The range between the

answers was taken into account to achieve a high degree of consensus about welfare measures and methods among experts. The best possible outcome is a trait or method of assessment, which combines a high US and a low range between the answers of the experts.

## RESULTS

### *Participating Experts*

In total, 35 of the 79 contacted experts responded to the invitation to join this study and completed ACA questionnaire 1. Subsequently, 11 out of these 35 experts were excluded; 9 because they did not have the minimum required years of experience and 2 due to a low consistency (correlation  $< 50\%$ ) in their answering. The 24 remaining experts finished their questionnaire in  $39 \pm 18$  min (mean  $\pm$  SD). The correlation coefficient within their answers was  $87.3 \pm 10.5\%$ . The 24 experts were from the following countries: UK, the Netherlands, Austria, Finland, and Italy (4 experts per country), and Germany, Switzerland, Ireland, and the Slovak Republic (1 expert per country). Six experts clarified that they had only 1 species in their field of experience, 11 had 2, and 7 had 3 species in their field of expertise, all including dairy cattle. An overview of the years of experience and number of publications as a first or co-author of these 24 experts is in Table 1.

From these 24 experts, 12 cooperated in the ACA questionnaires 2 and 3. Again, the 50% threshold value of the correlation coefficient was used as a selection criterion for consistency in the answering of the experts. From ACA questionnaire 2 on methods for assessing lameness cases, all experts exceeded this value ( $98.1 \pm 1.8\%$ ) and were included in the final results, in contrast with ACA questionnaire 3, where only 9 experts exceeded the 50% correlation coefficient (average  $76.4 \pm 16.6\%$ ). The average years of experience and number of publications as a first or co-author of these remaining experts of the second study and ACA questionnaire 3 are given in Table 1.

The most valuable measures within the category of high US are the ones with the lowest ( $\leq 15$ ) range of US (Table 2). Four measures met this criterion: lameness cases (US = 107.3), competition for feed and water (US = 96.4), number of freestalls per 10 cows (US = 84.8), and air turnover within the barn (US = 71.5), respectively. Stereotype behavior (US = 111.7 points), BCS (US = 108.0), and hock lesions (US = 104.7) were included in the high-US category, but with a medium (15 to 30) range of US (Table 2).

In the category of medium US (35 to 70), no measures had a low range of US (Table 3). The traits lying position in the freestalls (US = 69.7), humidity in the barn

**Table 1.** Selection criteria for the remaining experts in the first (n = 24), second (n = 12), and third (n = 9) Adaptive Conjoint Analysis (ACA; Sawtooth Software Inc., Sequim, WA) questionnaire

Inclusion criterion	Average	Minimum	Maximum
ACA questionnaire 1			
Years of experience	13.5	5	26
Publications as co-author in the last 5 yr	7.3	1	15
Publications as first author in the last 5 yr	3.6	1	7
ACA questionnaire 2			
Years of experience	8.7	5	14
Publications as co-author in the last 5 yr	4.0	1	8
Publications as first author in the last 5 yr	2.2	1	4
ACA questionnaire 3			
Years of experience	7.6	5	11
Publications as co-author in the last 5 yr	3.8	1	6
Publications as first author in the last 5 yr	2.1	1	3

(US = 68.4), and the form and size of the freestalls (US = 64.2) all yielded a medium range of US.

Because of the number of US with value <1 per trait in the category low US, often the medians of utility and range were equal (Table 4). Within this category the traits ratio of standing cows to total cows (US = 13.1), clipping of the cows (US = 12.2), and milk composition (US = 7.6) showed the highest US with a small range between the expert's answers. The items calf-cow contact, introduction of heifers, young stock rearing, and downer cow management were mentioned as missing measures during the questionnaire, but were only mentioned once per item. The highest scoring method in ACA questionnaire 2 on methods to assess lameness cases was findings of the hoof trimmer (US = 58.7; Table 5). Well-known methods such as the assessment according to Sprecher et al. (1997) or Manson and Leaver (1988) were rewarded with lower (US = 52.8) or no US at all, respectively. Assessment through activity measurements using a neck sensor or leg sensors, as well as through the finding of lesions by the farmer, were rewarded with zero points and, hence, classified as not useful.

The visual cleanliness of the floor and the freestall bedding (US = 67.3) were considered very important by respondents to ACA questionnaire 3 for assessing the hygiene of the calving pen (Table 6). It was the only method classified in the medium category of US; the other methods were classified in the low category.

## DISCUSSION

The ACA technique is a useful technique to rank measures for practically assessing the welfare in dairy herds. A large range in consistency of 76.4% between expert's answers for the methods of assessment was found, which did not meet a certain degree of international acknowledgment compared with 98.1% when the methods of assessment were well established. The de-

gree of international acknowledgment was based on the second selection criterion, published in double-refereed journals. The technique itself is not very common in animal (welfare) science and was only used once before by Angus et al. (2005). Because their study differs in the selected traits and the background of the selected experts in comparison to this study, this discussion section is divided into 3 parts: the results of the 3 ACA

**Table 2.** Ranking of the welfare measures in the high utility score (US) category ( $\geq 70$ ) sorted by their small, medium, or large range of US among the experts

Selected measure <sup>1</sup>	Median US <sup>2</sup>	Range <sup>3</sup>
Small range of US		
Lameness cases	107.3	11.7
Competition for feed and water	96.4	13.9
Number of freestalls per 10 cows	84.8	13.3
Air turnover within the barn	71.5	8.5
Medium range of US		
Stereotype behavior	111.7	17.1
BCS	108.0	18.9
Hock lesions	104.7	16.1
Locomotion (gait) score	96.6	22.7
Percentage of clinical mastitis	95.2	25.9
Aggression toward each other	88.5	25.9
Panting and sweating of the cows	87.0	19.9
Percentage of diarrhea	84.4	25.6
Separate calving pen	78.0	28.0
SCC in milk	76.3	22.3
Bedding of the freestalls	76.1	19.7
Hygiene of the calving pen	75.0	23.4
Large range of US		
Overcrowding	90.2	30.3
Flight distance	81.2	32.5
Percentage of clinical ketosis	80.2	31.2
Percentage of clinical milk fever	79.9	31.5
Cow cleanliness	75.5	37.7
Separate sick cow pen	74.5	34.7
Floor type	70.0	40.3

<sup>1</sup>Ranges in US among the experts: small range:  $\leq 15$ ; medium range: between 15 and 30; and large range:  $\geq 30$ .

<sup>2</sup>Median of all US per measure.

<sup>3</sup>Range in US among the experts.

**Table 3.** Ranking of the welfare measures in the medium utility score (US) category (35 to 70) sorted by their small, medium, or large range of US among the experts

Selected measure <sup>1</sup>	Median US <sup>2</sup>	Range <sup>3</sup>
Small range in US		
None		
Medium range in US		
Lying position in the freestall	69.7	16.4
Humidity in the barn	68.4	20.1
Form and size of the freestalls	64.2	24.9
Cleanliness of the floor	63.2	26.0
Average lying time	58.4	27.8
Hygiene of claws and hind legs	56.2	22.9
Temperature inside the barn	55.3	17.7
Position of the knee bar and neck rail	50.0	22.5
Self-grooming	45.6	23.5
Cleanliness of drinking bowls/feeding places	42.1	24.5
Frequency of fresh feed per 24 h	38.5	29.1
Cystic ovarian disease	36.4	27.4
Large range in US		
Voluntary animal approach	63.9	35.3
Caesarean section	62.6	33.2
Culling rate	58.4	46.9
Claw-trimming frequency	51.2	30.4
Dehorning	50.0	30.1
Vaginal discharge	39.9	30.9
Open side walls	38.7	36.2
Cow traffic at milking parlor	35.6	35.3

<sup>1</sup>Ranges in US among the experts: small range:  $\leq 15$ ; medium range: between 15 and 30; and large range:  $\geq 30$ .

<sup>2</sup>Median of all US per measure.

<sup>3</sup>Range in US among the experts.

questionnaires; the selection of the experts, traits, and methods of assessment; and the methodology itself.

For the results, the median utility and the absolute range of traits were taken into account. Other studies used only the mean as its ranking criteria and presumed an equal distribution of preferences among all measures. Because of a small number of participants in various studies, this assumption is questionable (Haslam and Kestin, 2003; Angus et al., 2005; Valeeva et al., 2005). To accomplish a standard protocol, the ranking must include a level of agreement among the experts instead of only ranking on the median US. This resulted in a lower ranking of the measures stereotypic behavior, hock lesions, and BCS, and a higher ranking of the traits number of freestalls per 10 cows and air turnover within the barn. These traits would otherwise have been ranked at the eleventh and twenty-second position instead of the third and fourth position (Table 2). Despite the higher US of overcrowding and percentage of clinical ketosis than air turnover within the barn, these traits are less useful because of their almost 4 times higher range of US between the experts.

Our data support the choices of 22 dairy cattle experts in a Delphi questionnaire (Whay et al., 2003a) for the traits prevalence of claw lesions, hock lesions, BCS, and percentage of clinical mastitis, diarrhea, ketosis, or milk fever. Other measures were classified much

lower: observed cow cleanliness and social behavior (including competition for feed and water), aggression towards each other, lying behavior, and flight distance. A different level of expertise or different (country or husbandry) backgrounds of the experts between the 2 studies could perhaps explain these differences. Prominent missing traits from the questionnaire of Whay et al. (2003a) compared with the current study were panting and sweating of the cows, average lying time, and stereotypic behavior.

A similar ACA study, which ranked traits of dairy cow welfare assessment, included veterinary practitioners experienced in the field of dairy cattle and veterinary students in their pre-graduation year (Angus et al., 2005). Only the opinions of experienced veterinary practitioners were reported. These veterinary practitioners gave the highest score to the measures access to forage, feed bunk space, BCS, foot conformation, walking into the parlor without encouragement, and hock lesions. The measure stereotypic behavior was categorized as a middle-ranking trait. Our highest-ranking trait, prevalence of claw lesions, was not very important, whereas underfoot conditions were very important. This could be the reason for the wide range in US of some measures in our ACA study, and confirms the need for a combination of US and its variation (i.e., range) between the experts. The threshold values used

**Table 4.** Ranking of the welfare measures in the low utility score (US) category ( $\leq 35$ ) sorted by their small, medium, or large range of US among the experts

Selected measure <sup>1</sup>	Median US <sup>2</sup>	Range <sup>3</sup>
Small range in US		
Ratio of standing cows to total cows	13.1	13.1
Clipping of the cows	12.2	12.2
Milk composition (ratio fat/protein)	7.6	7.6
Natural mating or AI	1.3	1.3
Using a footbath	1.1	1.1
Branding	0.1	0.1
Frequency of defecation	0.0	0.0
Vocal expression	0.0	0.0
Medium range in US		
Frequency of eating and drinking	34.8	17.6
Barn type	33.9	24.9
Hygiene of the freestalls	30.8	22.9
Rumination pattern	29.1	24.1
Feeding method	28.1	26.3
Mechanical ventilation	26.6	18.8
Hair coat	26.0	26.0
Ratio of lying cows to total cows	25.2	22.2
Kicking in milking parlor	20.4	20.4
Undigested fraction of the manure (score)	20.2	19.8
Mounting	19.2	19.2
Average milking time of the herd	18.9	18.5
Barn isolation	18.0	16.2
% heat detection within 60 d after calving	16.2	16.2
Size of the herd	15.8	15.8
Rumen fill score	15.7	16.0
Large range in US		
Brushes present	34.8	34.8
Exploration behavior	33.5	33.5
Color and consistency of the manure	32.7	32.7

<sup>1</sup>Ranges in US among the experts: small range:  $\leq 15$ ; medium range: between 15 and 30; and large range:  $\geq 30$ .

<sup>2</sup>Median of all US per measure.

<sup>3</sup>Range in US among the experts.

in this study were arbitrary and could influence the ranking of measures. For example, if in the high-utility category (Table 2) a range of 20 between experts indicated a low range, some traits would move to the top-ranking measures, but the traits with a large range

between the experts still continue to be low. Overall, including the range of US between the experts provides an additional criterion and a more solid, robust background in selecting traits for the assessment of welfare in dairy herds.

**Table 5.** The measures from Adaptive Conjoint Analysis (ACA; Sawtooth Software Inc., Sequim, WA) questionnaire 2 to assess lameness cases [ranked by the median utility score (US) and the respective range among the experts]

Selected measure <sup>1</sup>	Median US <sup>2</sup>	Range <sup>3</sup>
High US		
None	—	—
Medium US and medium range		
Findings of the hoof trimmer	58.7	23.7
Assessment according to Whay et al. (2003b)	36.4	25.9
Medium US and large range		
Only lameness cases recorded by the hoof trimmer	56.0	42.9
Assessment according to Sprecher et al. (1997)	52.8	49.4
Low US and low range		
Lameness cases recorded by the farmer	7.7	5.3
Assessment according to Manson and Leaver (1988)	0.0	0.0
Activity measurements of a neck sensor	0.0	0.0
Activity measurements of a leg sensor	0.0	0.0
Pathological findings of the farmer	0.0	0.0

<sup>1</sup>Ranges in US among the experts: small range:  $\leq 15$ ; medium range: between 15 and 30; and large range:  $\geq 30$ .

<sup>2</sup>Median of all US per measure.

<sup>3</sup>Range in US among the experts.

**Table 6.** The measures from Adaptive Conjoint Analysis (ACA; Sawtooth Software Inc., Sequim, WA) questionnaire 3 to assess the hygiene of the calving pen [ranked by the median utility score (US) and the respective range among the experts]

Selected measure <sup>1</sup>	Median US <sup>2</sup>	Range <sup>3</sup>
High US		
None		
Medium US and medium range		
Visual cleanliness of the floor and freestall bedding	67.3	8.2
Low US and small range		
Visual cleanliness of the water bowls	32.0	13.8
Visual cleanliness of the feed bunk	28.1	5.2
Visual cleanliness of the gear	24.0	8.9
Number of parities before cleaning the pen	8.4	8.4
Cleanliness of the farmer (boots)	4.6	4.6
Visual cleanliness of the walls	0.0	0.0
Low US and medium range		
Method of cleaning and disinfection	20.8	18.5

<sup>1</sup>Ranges in US among the experts: small range:  $\leq 15$ ; medium range: between 15 and 30; and large range:  $\geq 30$ .

<sup>2</sup>Median of all US per measure.

<sup>3</sup>Range in US among the experts.

The results from ACA questionnaires 2 and 3 could not be compared with other studies, but it is surprising that the findings of the hoof trimmer in ACA questionnaire 2 were preferred over the assessment methods of gait scoring the cows. One would expect a higher preference for the frequently used and easily performed method of gait scoring instead of data from the less frequent (once or twice a year) farm visits of a professional hoof trimmer.

As expected, a difference between the outcomes of ACA questionnaires 2 and 3 was the variation in the consistencies of the expert's answers. The existing internationally acknowledged methods for assessing lameness cases could explain the small variation in consistency in the answers of the experts in contrast to hygiene of the calving pen without approved methods of assessment. This could imply that, if the methods of assessment of certain measures were less internationally acknowledged, the usefulness of these measures in practice is questionable and more research is needed on what the best welfare indicators are in such systems. In other comparable studies, selection criteria of experts either are absent or not clearly defined (Whay et al., 2003a; Angus et al., 2005). This makes it difficult to compare the results of the different studies. The number of experts in this study who passed all selection criteria was comparable to the numbers reported by other animal science studies that used the ACA method, which vary from 5 to 52 (van Schaik et al., 1998; Haslam and Kestin, 2003; Valeeva et al., 2005).

A disadvantage of the ACA technique is that once a questionnaire is composed and sent to the participants, no possibility exists to change it or add missing traits. A comparable method of ranking preferences of experts, the Delphi Technique (Linstone and Turoff,

2002), can overcome these problems, but deals less satisfactorily with inconsistent answers and bias. The major advantages of the ACA technique are rapid and easy access to experts, the objective ranking of the presented items, the objective way it measures consistency between the answers, the limited duration of answering the respective questions (about 1 h in our study), and the relatively limited costs involved. Consistency is especially important when the subject being investigated is hard to define through objective measurements, such as hygiene (in ACA questionnaire 3 on the assessment of hygiene of the calving pen). Using this technique, it was possible to exclude from the final calculations those experts who did not understand the rationale behind the questionnaire or changed their preferences during the questionnaire. Consequently, 25% of the experts used for ACA questionnaire 3 (assessment of hygiene of the calving pen) had to be excluded from the final calculations.

## CONCLUSIONS

As demonstrated in this study, the ACA technique is a useful technique to rank preferences of experts by means of an objective measure of consistency in their answers. The method was useful for both well-established and inconclusive methods of welfare assessment. Although the number of experts and the extent of their geographical distribution were limited, the applied method provided insight in which measures already available on European dairy farms were the most important ones to focus on in future research and to include in the phase of developing a welfare assessment protocol. From the highest ranking traits (US  $\geq 70$ ) it was possible to identify 3 major areas of interest.

The most important is lameness (measures: lameness case, hock lesions, and locomotion score), followed by a suitable environment (measures: number of freestalls, feeding places, climate, and panting and sweating), and lastly, a proper monitoring of endemic diseases (measures: clinical mastitis, SCC, and diarrhea). Future research should focus on including these 3 areas into a protocol that is supportive to and feasible for the dairy industry itself and transparent for consumers. In addition, the method that was applied to rank suitable welfare traits should be tested in other parts of the world beside the European Union to assess validity.

### ACKNOWLEDGMENTS

The authors express their gratitude to all participating experts for their time and effort devoted to filling in and returning the adaptive conjoint questionnaires. Also, many thanks to the ICT department of the Faculty of Veterinary Medicine, Utrecht University, the Netherlands, for providing the knowledge and assistance for transferring the adaptive conjoint program software to the World Wide Web.

### REFERENCES

- Angus, L. J., H. Bowen, L. A. S. Gill, T. G. Knowles, and A. Butterworth. 2005. The use of conjoint analysis to determine the importance of factors that affect on-farm welfare of the dairy cow. *Anim. Welf.* 14:203–213.
- Bartussek, H. 1999. A review of the animal needs index (ANI) for the assessment of animals' well-being in the housing systems for Austrian proprietary products and legislation. *Livest. Prod. Sci.* 61:179–192.
- Blokhuis, H. J., R. B. Jones, I. Veisser, and R. Geers. 2006. Measuring and Monitoring Farm Animal Health. K. U. Leuven R&D–Zoo-technical Centre, Lovenjoel, Belgium.
- Bracke, M. B. M., B. M. Spruijt, J. H. M. Metz, and W. G. P. Schouten. 2002. Decision support system for overall welfare assessment in pregnant sows A: Model structure and weighting procedure. *J. Anim. Sci.* 80:1819–1834.
- Capdeville, J., and I. Veissier. 2001. A method of assessing welfare in loose housed dairy cows at farm level, focusing on animal observations. *Acta Agric. Scand. A* 51:62–68.
- European Commission. 2006. Animal welfare main community legislative references. Commission Decision (2006/778/EC). Accessed February 2011. [http://ec.europa.eu/food/animal/welfare/references\\_en.htm](http://ec.europa.eu/food/animal/welfare/references_en.htm).
- FAWC (Farm Animal Welfare Council). 1992. Farm Animal Welfare Council: FAWC updates the Five Freedoms. *Vet. Rec.* 131:357.
- Frewer, L. J., A. Kole, S. M. A. Van de Kroon, and C. de Lauwere. 2005. Consumer attitudes towards the development of animal-friendly husbandry systems. *J. Agric. Environ. Ethics* 18:345–367.
- Haslam, S. M., and S. C. Kestin. 2003. Use of conjoint analysis to weight welfare assessment measures for broiler chickens in UK husbandry systems. *Anim. Welf.* 12:669–675.
- Linstone, H. A., and M. Turoff. 2002. *The Delphi Method: Techniques and applications*. Accessed July 2010. <http://www.is.njit.edu/pubs/delphibook>.
- Main, D. C. J., F. Webster, and L. E. Green. 2001. Animal welfare assessment in farm assurance schemes. *Acta Agric. Scand. A Anim. Sci.* 51:108–113.
- Main, D. C. J., H. R. Whay, L. E. Green, and A. J. F. Webster. 2003. Preliminary investigation into the use of expert opinion to compare the overall welfare of dairy cattle farms in different farm assurance schemes. *Anim. Welf.* 12:565–569.
- Manson, J. F., and J. D. Leaver. 1988. The influence of concentrate amount on locomotion and clinical lameness in dairy cattle. *Anim. Prod.* 47:185–190.
- Sawtooth Software Inc. 2005. *Adaptive Conjoint Analysis User Manual. Version 5.2*. Sawtooth Software Inc., Sequim, WA.
- Sprecher, D. J., D. E. Hostetler, and J. B. Kaneene. 1997. A lameness scoring system that uses posture and gait to predict dairy cattle reproductive performance. *Theriogenology* 47:1179–1187.
- Valeeva, N. I., M. P. M. Meuwissen, A. G. J. M. Oude Lansink, and R. B. M. Huirne. 2005. Improving food safety within the dairy chain: An application of conjoint analysis. *J. Dairy Sci.* 88:1601–1612.
- van Schaik, G., A. A. Dijkhuizen, R. B. M. Huirne, and G. Benedictus. 1998. Adaptive conjoint analysis to determine perceived risk factors of farmers, veterinarians and AI technicians for introduction of BHV1 to dairy farms. *Prev. Vet. Med.* 37:101–112.
- Webster, A. J. F. 2001. Farm animal welfare: The five freedoms and the free market. *Vet. J.* 161:229–237.
- Webster, A. J. F. 2003. Assessment of animal welfare at farm and group level: Introduction and overview. *Anim. Welf.* 12:429–431.
- Webster, J. 2005. The assessment and implementation of animal welfare: theory into practice. *Rev. Sci. Tech.* 24:723–734.
- Whay, H. R., D. C. J. Main, L. E. Green, and A. J. F. Webster. 2003a. Animal-based measures for the assessment of welfare state of dairy cattle, pigs and laying hens: Consensus of expert opinion. *Anim. Welf.* 12:205–217.
- Whay, H. R., D. C. J. Main, L. E. Green, and A. J. F. Webster. 2003b. Assessment of the welfare of dairy cattle using animal-based measurements: Direct observations and investigation of farm records. *Vet. Rec.* 153:197–202.
- Winckler, C., J. Capdeville, G. Gebresenbet, B. Hörning, U. Roiha, M. Tosi, and S. Waiblinger. 2003. Selection of parameters for on-farm welfare-assessment protocols in cattle and buffalo. *Anim. Welfare* 12:619–624.