Animal Welfare 2009, 18: 417-425
ISSN 0962-7286

Validation of an Animal Needs Index for cattle using Test Theory

T Herva*†, O Peltoniemi‡ and A-M Virtala§

† Quality Manager, Atria Ltd, Department of Production Animal Medicine, University of Helsinki, Finland
‡ Department of Production Animal Medicine, University of Helsinki, Finland
§ Department of Basic Veterinary Sciences, University of Helsinki, Finland
* Contact for correspondence and requests for reprints: tuomas.herva@atria.fi

Abstract

The validity of an ANI-based welfare index for cattle, the A-Index, was evaluated using Test Theory methods described for quality of life assessments in human medicine. Content validity was considered during the modification of the Index for Finnish beef production. In total, 43 items evaluating locomotion abilities, lying area, social environment, management, feeding, and health of animals over six months up to slaughter were formed. Index scorings were performed on 237 farms. A most consistent partial A-Index was constructed based on qualitative item analyses. A positive relationship was established between the full A-Index and daily carcase gain. In addition, a preliminary negative relationship was established between the partial A-Index, body fat and mortality. Based on our observations, methods used for quality of life indices in human beings would appear suitable for animal welfare assessment. Practicability and well-defined methodology are the main advantages of Test Theory approach.

Keywords: *animal welfare, cattle, daily gain, epidemiology, fat, mortality*

Introduction

The tradition in animal welfare science has been to measure different behavioural, physiological and health parameters under experimental conditions (Bracke 2007). Botreau *et al* (2007a) described different ways to assess on-farm welfare. Data concerning a farm can be analysed by an expert who draws an overall conclusion. The data can also be compared with minimal requirements set for each measure, converted into ranks, and summed, or converted into values or scores compounded in a weighted sum (eg TGI35L). A final possibility is to use *ad hoc* rules. Summated scales are considered to be the most suitable for overall welfare assessment despite having been criticised for many reasons (Botreau *et al* 2007a). Indices mainly comprise environment-based measures, although animal-based measures, such as behavioural and health parameters, are generally considered to be more closely linked to the welfare of animals (Capdeville & Veissier 2001; Whay *et al* 2003; Winckler *et al* 2003). It has also been questioned if there is any rational way to define the weights of individual scores (Amon *et al* 2001). During recent years, the theoretical basis for overall assessment of animal welfare has been thoroughly discussed (Botreau *et al* 2007b).

Animal welfare can be conceptually compared with human quality of life. Neither is a directly measurable quantity, but rather a concept, which can be estimated by Test Theory methods described by Nunally and Bernstein (1994), and widely used in psychometrics, educational and social sciences. The concept of interest, such as the level of animal welfare, can be termed a latent variable. Although it cannot be measured directly, the magnitude of the latent variable can be assumed to vary in different situations. Methods to measure a latent variable by summated scales are well described (eg De Vellis 2003). The best set of items (questions) for a scale to measure a particular latent variable depends not only on the concept of interest but also the measured population (Nunnally & Bernstein 1994). For example, in the case of a mathematical test, a very easy set of variables does not differentiate skilled pupils at all and a difficult item does not measure mathematical skills if the pupils have seen the correct answers in advance. Similarly, we may ask, whether fear of humans is an important part of animal welfare in Finnish beef production. If animals were tame on most of the farms, fearfulness would not differentiate farms according to overall welfare status, although fear shown towards humans can represent a major welfare problem for animals.

Consequently, compared with traditional methods used in animal welfare sciences, Test Theory is a conceptually contradictory way to study animal welfare. In animal welfare sciences, as in life sciences in general, a phenomenon of interest (such as animal welfare) is first divided into directly measurable components (such as lying time, locomotion score or cortisol cycle). Conclusions concerning the phenomenon of interest are drawn from these measurements, usually without using any formalised methods. In contrast, Test Theory uses measures based on the theory concerning the phenomenon of interest. The aim of the Test

Theory is not to build an ultimate tool to measure certain phenomena, but rather to provide a formalised method to study phenomena in certain populations or situations in a practical manner. For example, there is no need to build one mathematical test suitable for every situation, but rather to know how to build the best test for a certain mathematical course. Accordingly, Test Theory could enable measurement of animal welfare as a single phenomenon in a well-defined manner. A continuous summated scale, a welfare index, could also make it easier to compare farms and study the associations of welfare and other farm parameters.

On-farm animal welfare levels have been regularly evaluated in Austria and Germany using an Animal Needs Index (ANI) (Bartussek 1999) The ANI was modified for meat production animals in a Finnish meat company, Atria Ltd and adopted as the A-index to be part of a meat quality programme (Munsterhjelm & Herva 2003). The main aim of the present study was to validate the A-Index as an overall welfare measure using Test Theory methods.

Materials and methods

Coverage (ie content validity), reliability, responsiveness (ie construct validity) and sensitivity of the A-Index for measuring the on-farm welfare of bulls between 6 and 24 months of age were evaluated using criteria developed for the quality of life assessments for human beings (Testa & Simonson 1996). The concepts are used in the present article according to Test Theory (Nunnally & Bernstein 1994). Consequently, the concepts differ partly from those applicable to the conventional context of life sciences.

In the context of Test Theory, content validity refers to a scale's ability to cover all aspects of the latent variable of interest (DeVellis 2003). Reliability is defined as the proportion of a scale variance attributable to the true score of the latent variable. It is widely measured using Cronbach's alpha which evaluates internal consistency of a scale. However, it can also be measured by using a parallel test or using test-retest reliability. In this sense, reliability resembles repeatability used in life sciences. Construct validity is a measure of the association between the change in the observed summated scale, and the change in the true value of the latent variable. Because it cannot be measured directly, it is often assessed using a criterion variable, which is causally associated with the latent variable. In the context of Test Theory, sensitivity refers to the ability of a scale to reflect true changes in the latent variable.

In this study, animal welfare was regarded as a prolonged mental state, resulting from how the animal experiences its environment over time (Dawkins 1980; Duncan 1996; Bracke *et al* 1999). Inclusion of items needed to ensure content validity of the A-Index as a welfare indicator was considered during the development process by a farm advisory group of Atria Ltd. The construction of the A-Index was based on an ethical justification process using the reflective equilibrium method (DeGrazia 1996). In this process, each item was discussed by the group until consensus concerning weights and formulations for different values of the item was reached. Modifications of the ANI were based on the literature and practical plausibility. Applicability in the local commercial production environment was the main development criterion. In total, 43 items, scoring management, production environment, feeding and health of the animals were included in the A-Index, with maximal score 100, to assure content validity (Scott *et al* 2001). Used references were included in the A-Index (Table 1). Best and worst level for each item were reported previously (Herva *et al* 2009). There were minimal requirements set for certain items to be used in a farm quality programme. Those limits were not included in the analyses of this study.

Scorings were carried out by the advisory personnel of Atria Ltd on 237 farms in 2003. Each farm was scored only once. Items were evaluated based on these scorings using widely-accepted scale development criteria (DeVellis 2003). Item difficulty, correlation between each item and sum of the scale (Item test correlation), correlation between each item and sum of the scale excluding the particular item (Item rest correlation, *r*) and Cronbach's alpha are commonly-used parameters in item analysis. Difficulties for each item were calculated by dividing the mean value of the item by the maximum value and subtracting the quotient from 1. Cronbach's alpha item test and item rest correlations were computed using the Intercooled Stata 9.0 programme for Windows (StataCorporation, College Station, TX, USA 2005).

Factor analysis, principal component analyses and item response analyses are regularly used methods in Test Theory, but not suitable tools to evaluate the A-Index because this index included items with only two alternatives and scale length varied among items. Subsequently, the ability of each item to discriminate farms with good welfare (high A-Index score) from those with poor welfare (low A-Index score) was evaluated qualitatively, comparing item test correlation with item difficulty (Figure 1). Difficult (difficulty near 1) items with item rest correlation over 0.2 were recognised to discriminate farms with excellent welfare from farms with good welfare. Items with moderate difficulty (near 0.5) discriminated average farms from each other and easy items (difficulty under 0.4) were used for discrimination among farms with poor welfare. Very easy items (difficulty near 0) did not differentiate the farms because almost all farms received the maximum score in these items regardless the farm's A-Index score. These items had also a low item test correlation. An item was considered to be 'poor' if it was very easy, negatively correlated with the A-Index score or did not occur consistently with the other items (item rest correlation < 0.2). In Test Theory, there are not any ultimate inclusion criteria, although item test correlation 0.3 is referred to as the arbitrary cut-off point for item inclusion (Nunnally & Bernstein 1994). We used an item rest correlation of < 0.2 for inclusion criterion to avoid the inflating effect of the particular item. Consequently, item test correlation of selected items remained over 0.28 (Figure 1).

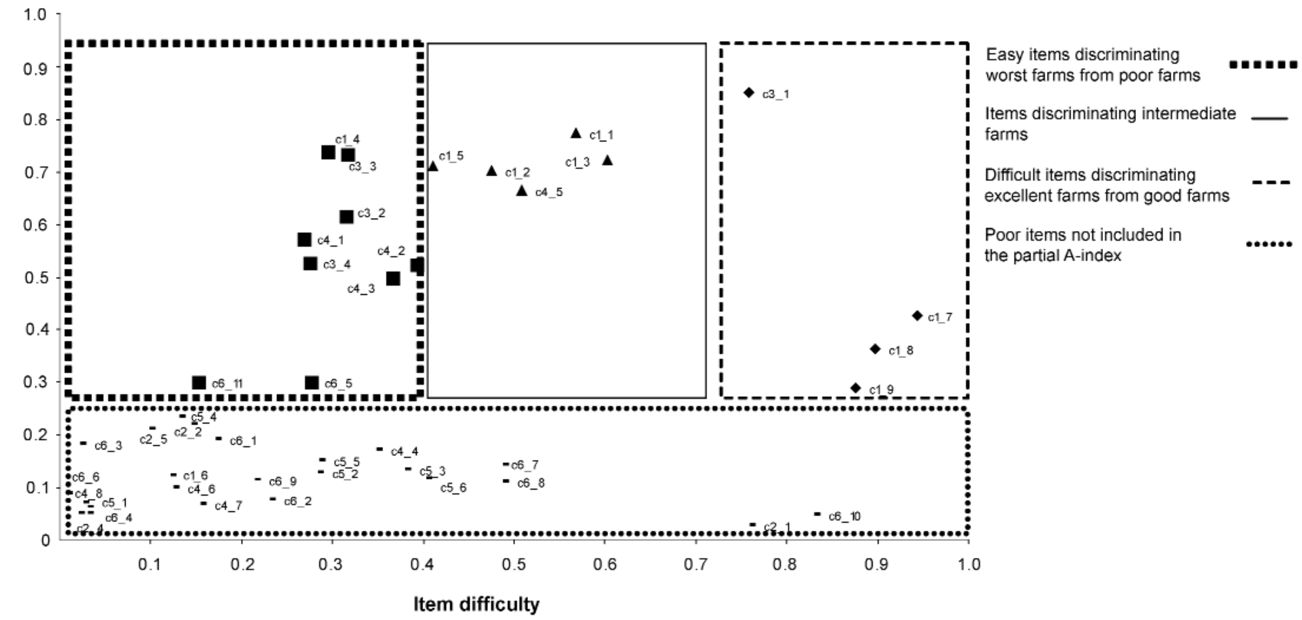
A most consistent summated subscale, termed the partial A-Index, was built from 'good' items correlating with total score and was used as a general welfare indicator. 'Poor'

Table 1 Item description of an on-farm welfare index (A-Index) for cattle from 6–24 months of age. Mean score and item rest correlation is based on A-Index scorings on 237 farms in Finland in 2003.

Item no	Item name	Worst level used as minimum requirement in meat quality programme of Atria Ltd	Max score	Mean score	Item rest correlation to full score excluding the item	Included in the Partial A-Index
c1_1	Freedom of movement, pen shape (Jóhannesson & Sørensen 2000)	N	2	0.87	0.74	Y
c1_2	Freedom of movement, confrontations	N	3	1.58	0.65	Y
c1_3	Space allowance (Andersen *et al* 1997; Ingvartsen & Andersen 1993)	Y	3	1.19	0.68	Y
c1_4	Lying down and getting up possibilities	Y	3	2.11	0.71	Y
c1_5	Floor quality on walking and feeding area	N	2	1.18	0.69	Y
c1_6	Condition of pen fixtures	N	1	0.88	0.11	N
c1_7	Outdoor access in winter	N	2	0.11	0.40	Y
c1_8	Grazing/outdoor access in summer	N	2	0.21	0.32	Y
c1_9	Outdoor enclosures	N	2	0.25	0.24	Y
c2_1	Dehorning	N	2	0.48	–0.29	N
c2_2	Regrouping of animals	N	3	2.56	0.15	N
c2_3	Group size (Menke *et al* 1999)	N	2	1.66	–0.32	N
c2_4	Human-animal relationship	N	2	1.94	0.03	N
c2_5	Feeding space allowance (Andersen *et al* 1997)	N	2	1.80	–0.18	N
c3_1	Softness of resting area (Hannan & Murphy 1983)	N	6	1.45	0.77	Y
c3_2	Clean resting area (Ruis-Heutinck *et al* 2000)	N	6	4.11	0.52	Y
c3_3	Non-slippery resting area	N	2	1.37	0.71	Y
c3_4	Dry resting area	N	2	1.45	0.49	Y
c4_1	Indoor air quality and ventilation capacity	Y	4	2.93	0.51	Y
c4_2	Draught on resting area	N	4	2.43	0.46	Y
c4_3	Water availability (Andersson 1984)	N	4	2.53	0.42	Y
c4_4	Water temperature (Andersson 1984)	N	2	1.30	0.13	N
c4_5	Noise	Y	2	0.98	0.62	Y
c4_6	Light regime (Phillips *et al* 2000)	N	2	1.75	0.06	N
c4_7	Night light (Phillips *et al* 2000)	N	1	0.84	0.03	N
c4_8	Day light (Phillips *et al* 2000)	Y	1	0.99	0.08	N
c5_1	Roughage availability (Lindström & Redbo 2000)	N	2	1.94	0.04	N
c5_2	Roughage quality	N	2	1.43	–0.08	N
c5_3	Concentration feeding method	N	3	1.86	–0.06	N
c5_4	Concentration ratio	N	3	2.60	–0.17	N
c5_5	Ration formulation	N	2	1.42	0.08	N
c5_6	Daily carcase gain preceding scoring	N	3	1.79	0.04	N
c6_1	Body hygiene of animals	Y	3	2.48	0.15	N
c6_2	Hygiene of feeding and watering equipment	N	2	1.54	–0.03	N
c6_3	Injuries caused by pen fixtures (Schrader *et al* 2001)	N	1	0.98	0.17	N
c6_4	Skin condition (Menke *et al* 1999)	N	2	1.96	0.03	N
c6_5	Foot and leg injuries and diseases (Hassall *et al* 1993; Ruis-Heutinck *et al 2000)*	N	2	1.45	0.25	Y
c6_6	Other disease	N	1	0.97	0.06	N
c6_7	Disease recording	Y	1	0.51	0.12	N
c6_8	Sick pen	N	3	1.54	–0.02	N
c6_9	Euthanasia	N	1	0.78	–0.09	N
c6_10	Handling facilities	N	1	0.17	0.01	N
c6_11	Loading facilities	Y	1	0.85	0.27	Y
FAIc	**Full A-index**		**100**	**63**		
PAIc	**Partial A-Index**		**49**	**27**		

https://doi.org/10.1017/S0962728600000828 Published online by Cambridge University Press

Figure 1



Qualitative item analysis of the A-Index for cattle from 6 up to 24 months of age using data from 237 farms scored in 2003. Items with a negative item test correlation not included in the figure. Codes signify the following: c1_1 = Freedom of movement, pen shape; c1_2 = Freedom of movement, confrontations; c1_3 = Space allowance; c1_4 = Lying down and getting up possibilities; c1_5 = Floor quality on walking and feeding area; c1_6 = Condition of pen fixtures; c1_7 = Outdoor access in winter; c1_8 = Grazing/outdoor access in summer; c1_9 = Outdoor enclosures; c2_1 = Dehorning; c2_2 = Regrouping of animals; c2_3 = Group size; c2_4 = Human-animal relationship; c2_5 = Feeding space allowance; c3_1 = Softness of resting area; c3_2 = Clean resting area; c3_3 = Non-slippery resting area; c3_4 = Dry resting area; c4_1 = Indoor air quality and ventilation capacity; c4_2 = Draught on resting area; c4_3 = Water availability; c4_4 = Water temperature; c4_5 = Noise; c4_6 = Light regime; c4_7 = Night light; c4_8 = Day light; c5_1 = Roughage availability; c5_2 = Roughage quality; c5_3 = Concentrate feeding method; c5_4 = Concentrate ratio; c5_5 = Ration formulation; c5_6 = Daily carcase gain preceding scoring; c6_1 = Body hygiene of animals; c6_2 = Hygiene of feeding and watering equipment; c6_3 = Injuries caused by pen fixtures; c6_4 = Skin condition; c6_5 = Foot and leg injuries and diseases; c6_6 = Other disease; c6_7 = Disease recording; c6_8 = Sick pen; c6_9 = Euthanasia; c6_10 = Handling facilities; c6_11 = Loading facilities.

items were removed because they were not regarded as good welfare indicators for the studied population. Cronbach's alphas for the full A-Index and the partial A-Index, as well as the effect of each item on the alphas, were determined to estimate reliability of the scales. Cronbach's alpha is derived from the number of items used in a scale (N), sum of variance of each item ($\Sigma_{i=1} \sigma_{Yi}^2$) and variance of summated scale (σx^2):

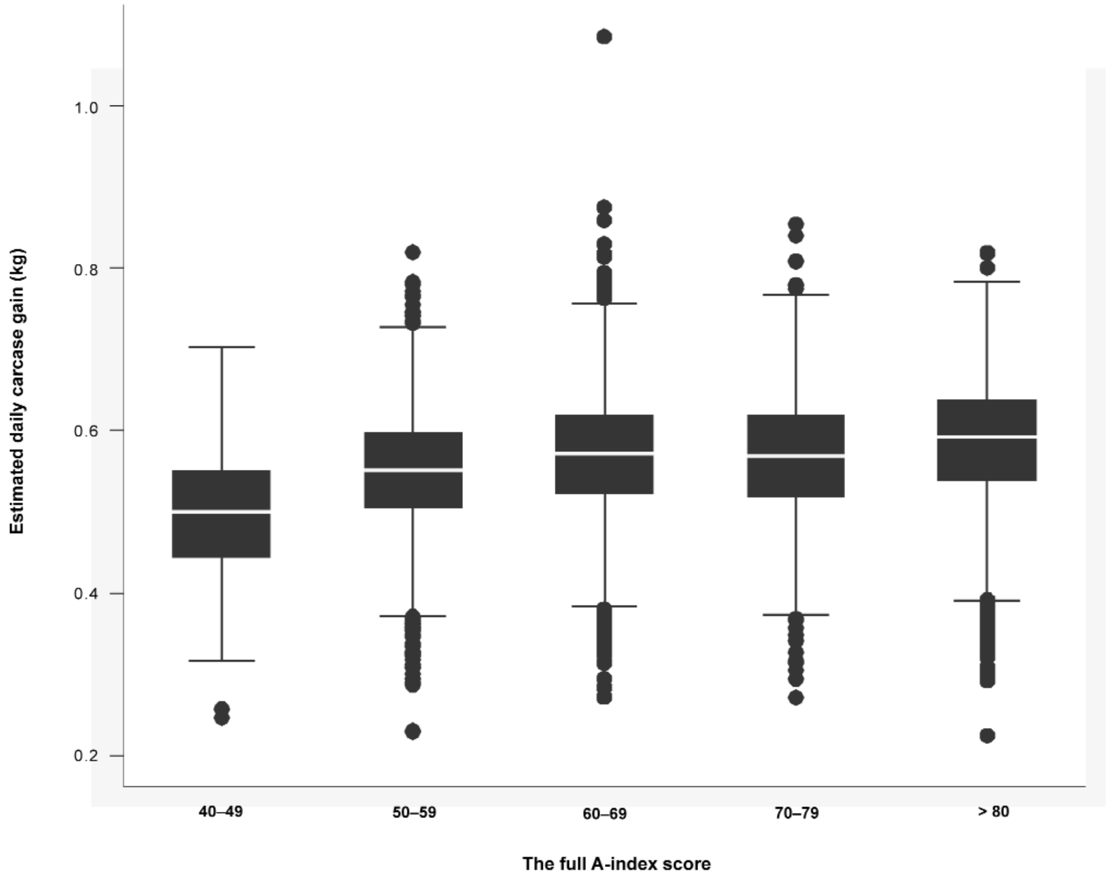
$$\alpha = N/(N-1) \times (1 - \Sigma_{i=1} \sigma_{Yi}^2/\sigma x^2)$$ (Cronbach 1951).

Construction validity and sensitivity of the index were evaluated based on the relationship between measured on-farm welfare and daily carcase gain, and body fat at slaughter and mortality from six months up to slaughter. According to the reviewed literature, it was supposed that on-farm welfare is positively associated with daily carcase gain and negatively associated with fat score at slaughter and mortality (Webster *et al* 1972; Broom 1991). Poor growth can be a sign of decreased welfare. However, fast growth is not a guarantee of good welfare (Broom 1991). Fat score at slaughter can be compared with body condition scoring, which is a widely-used welfare parameter (Rushen 2003). Fat scores varied between 1 and 5, where a score of 1 may indicate starvation, insufficient feeding or potential health problems. However, we were more interested in studying whether increased prevalence of scores 3, 4 or 5 could be associated with decreased welfare. A well-known decreasing effect of poor welfare and stress on growth rate is hypothesised to be due to increased proportion of energy retained as fat (Webster *et al* 1972). Stress in rats and humans has a well-known increasing effect on abdominal fat, which is connected to type II diabetes, increased cardiovascular morbidity, and mortality (Dallman *et al* 2003). Reduced life expectancy indicates that the animal has been stressed and that its welfare, at some point or points during its life, has been poor (Broom 1991).

Studies concerning these associations will be described in detail in separate articles with a very limited degree of overlapping data (Herva *et al* 2009; Herva *et al* unpublished data). Confounding effects of clustering by farms and a range of variables derived from animal delivery and slaughter data were checked for using mixed models (Dohoo *et al* 2003). Historical cohorts of animals under certain

https://doi.org/10.1017/S0962728600000828 Published online by Cambridge University Press

Figure 2



Box-and-whisker plot showing the relationship between on-farm welfare, measured using the full A-Index and, estimated daily carcase gain of 13,787 slaughtered bulls delivered first time in 2003 by Atria Ltd.

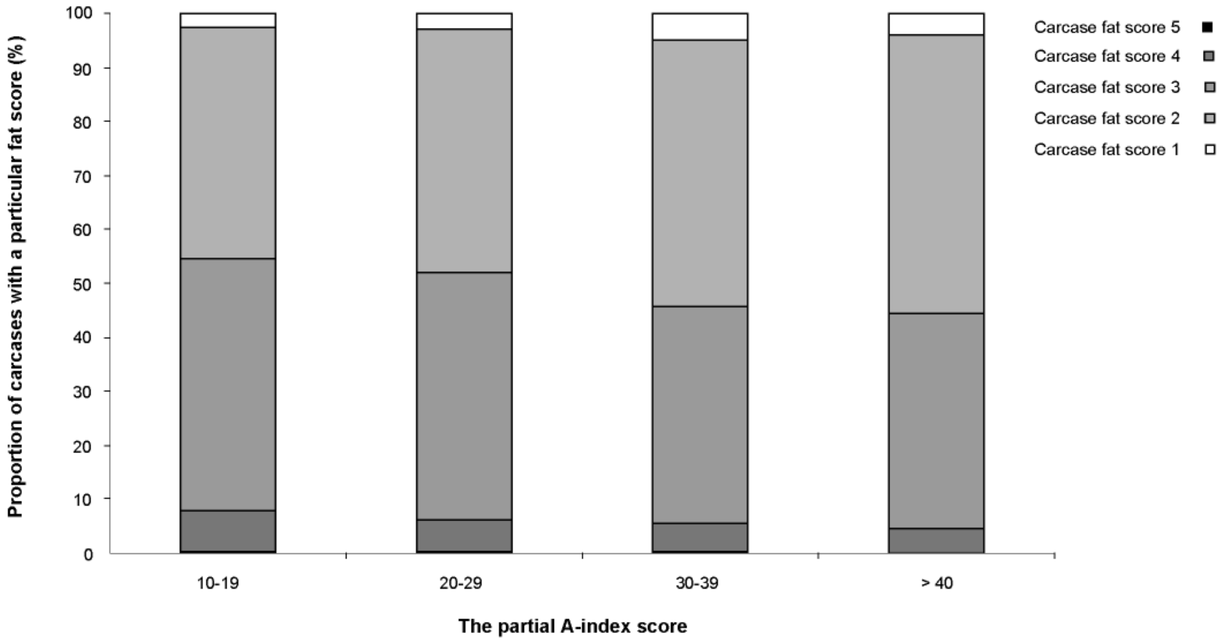
welfare status were formed. For this purpose, a total of 55,375 calves delivered by Atria Ltd for the first time in 2003 were selected for the study. Size of the first delivery farm and the finishing farm were estimated from the total numbers of calves delivered to the particular farms. Animals at a certain age were associated with the A-Index scorings of the corresponding age group using the identity of the farm where animals had been under a particular age. For 13,738 of the 55,375 calves delivered in 2003, scoring results for animals from 6 months up to 24 months of age were available from 168 farms. Carcase weight and fat score were recorded at slaughter, according to official standards and regulations used in the European Union (Council Regulation [EEC] No. 1208/81). The slaughter information was collected from the slaughterhouse database. An estimated daily carcase gain was calculated by extracting estimated birth weight of 16 kg from carcase weight and dividing the difference by age at slaughter. Date and reason for culling or death of non-slaughtered animals were collected from the National Animal Identification Register for Cattle. Death and euthanised calves were included in mortality.

Results

After evaluation, the 18 items shown in Table 1 were included in the most consistent partial A-Index. The full A-Index from 237 farms for young stock ranged from 38 up to 89.5. The median was 61.5 and mean 63. Maximum score of the partial A-Index was 49, mean 27, median 24 and minimum 11. Mean item difficulty in the full A-Index was 0.34, median 0.29, maximum 0.94 and minimum 0.01. Item difficulty in the partial A-Index ranged from 0.15 up to 0.90. The item evaluating group size was excluded from the partial A-Index due to negative correlation with full score. Cronbach's alpha was 0.82 for the full A-Index and 0.87 for the partial A-Index.

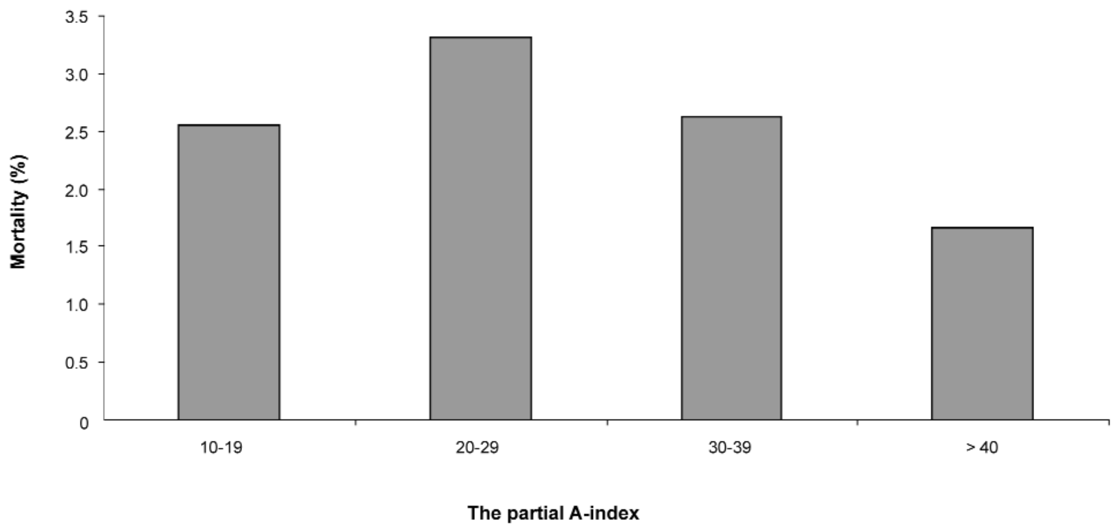
A positive one-to-one relationship (Pearson's correlation coefficient 0.19, $P < 0.0001$) was established between the full A-Index and estimated daily carcase gain (Figure 2). The relationship was confirmed using a mixed model (Herva *et al* 2009). A significant negative one-to-one relationship ($P < 0.0001$ for Pearson's chi-squared test) between the partial A-Index and carcase fat score

Figure 3



Distribution of carcase fat score of 12,675 slaughtered bulls delivered first time in 2003 by Atria Ltd by on-farm welfare measured by the partial A-Index score. Carcase fat score (1–5) was recorded at slaughter according to official standards and regulations used in the European Union (Council Regulation [EEC] No. 1208/81).

Figure 4



Mortality of 12,976 bulls between 6 to 24 months of age by on-farm welfare measured using the partial A-Index. The bulls were delivered first time by Atria Ltd in 2003.

(Figure 3) was preliminarily confirmed using a mixed model (Herva *et al* unpublished data). Association between mortality of bulls and on-farm welfare (Figure 4) was investigated in a preliminary study (Herva *et al* unpublished data). Mortality of bulls from 6 months of age up to slaughter was 2.6% with the partial A-Index score between 10 and 19, 3.3% with a score of 20 and 29, 2.6% with a score of 30–39, and 1.7% with a score of 40 and over (P = 0.001 for Pearson's chi-squared test). Moderate mortality in animals with the lowest partial A-Index score seemed to be explained in preliminary mixed models by smaller group size for these animals.

Discussion

Based on our observations, the methods used to develop quality of life indices for human beings seemed to be appropriate and practical to use for on-farm animal welfare assessment. The partial A-Index was a consistent indicator of general on-farm welfare, whereas the full A-Index was more appropriate as a tool to evaluate the production environment on farms.

Content validity of the full A-Index was considered to be good for comparing selected items with novel principles for welfare assessment. A set of 12 principles has been proposed to develop systems for welfare monitoring in the Welfare Quality® project: absence of prolonged hunger; absence of prolonged thirst; comfort around resting; thermal comfort; ease of movement; absence of injuries; absence of disease; absence of pain induced by management procedures; expression of social behaviour; expression of other behaviour; good human-animal relationship and absence of general fear (Botreau *et al* 2007b). All of these principles were covered at least to some extent in the full A-Index score. However, there has been substantial progress in evaluation of animal-based welfare since 2002 (Winckler *et al* 2003) when the A-index was developed.

The most suitable scale to evaluate a latent variable depends on the tested population (Nunnally & Bernstein 1994). Although an item might markedly affect animal welfare, it can be an inappropriate part of a summated scale to be used as an overall welfare measure for a studied population if it does not occur consistently with other indicators or it is not stringent enough to differentiate farms.

Items evaluating a good human-animal relationship, absence of prolonged hunger, general fear and diseases other than leg problems were not included in the partial A-Index due to poor reliability. Most of the rejected items were not stringent enough (difficulty near 1) to provide any extra information for the partial A-Index. Items scoring dehorning, water temperature, roughage quality, concentrate feeding method, ration formulation, daily carcase gain preceding scoring, feeding and watering equipment hygiene, disease recording, sick pens, euthanasia, and handling facilities were not consistent with other items due to low correlation with the A-Index score. It is arguable whether these principles are an important part of overall welfare of studied bulls from 6 to 24 months of age. Leg problems were the only animal-based parameter included in the partial A-Index. In this respect, the content validity of the partial A-Index was only reasonable. It can also be questioned whether it is possible to develop animal-based parameters occurring more consistently with the overall welfare.

Overall reliability of the full A-Index and the partial A-Index can be considered to be good as a result of high Cronbach's alphas, 0.82 and 0.87, respectively. Regarding internal consistency of a scale, an alpha below 0.6 is unacceptable, up to 0.65, undesirable, between 0.65 and 0.7, minimally acceptable, between 0.7 and 0.8, respectable and between 0.8 and 0.9, very good (DeVellis 2003). If alphas are much above 0.9, one should consider shortening the scale. There are two previous reports concerning reliability using internal consistency of on-farm welfare measurements. On pig farms (Munsterhjelm *et al* 2006), Cronbach's alphas for an index were 0.66 for farrowing, 0.84 for breeding and 0.91 for gestating sow units. For the different subscales, alpha ranged from 0.15 to 0.91. Subscales meeting the suggested 0.60 limit were 'locomotion' in all units (farrowing, breeding and gestation) as well as 'floor quality', 'climate', 'feeding', and 'health and stockmanship' in gestation. In a welfare index for horses (Beyer 1998), Cronbach's alpha for 'feeding equipment and resting area' was 0.31, for 'constructions,' 0.87, for 'group management,' 0.48, for 'outdoor access,' 0.48, for 'health control,' 0.80, and for 'enclosure management,' 0.96. In a study using conventional methods to measure repeatability in life sciences with multiple assessors, the repeatability was 0.93 for 'freedom to move', 0.89 for 'social contact', 0.70 for 'floor design', 0.75 for 'light-air-noise', 0.20 for 'looking after the animals' and 0.94 for total TGI score (Amon *et al* 2001). In our study, the reliability for items scoring resting area was better and the reliability for items scoring social contacts worse compared to repeatability gained with multiple assessors.

Reported poor reliability of indices comprising animal-based measures is a challenge for on-farm welfare assessment. Welfare is a prolonged mental state, resulting from how the animal experiences its environment over time (Dawkins 1980; Duncan 1996; Bracke *et al* 1999). Resource-based methods used to measure overall on-farm welfare have been widely criticised and animal-based methods have been emphasised (Keeling 2005). Repeatability of suitable welfare parameters (Winckler *et al* 2003) and strategy (Botreau *et al* 2007b) for overall welfare assessment tools are described. However, suitability of these parameters for overall assessment scales and internal consistency of potential scales remains unclear.

Our study design is in line with suggested principles of on-farm assessment schemes based on an epidemiological approach (Waiblinger *et al* 2001). We studied factors affecting welfare through the A-Index. Instead of the several health and behavioural parameters suggested (skin lesions; prevalence of leg disorders and mastitis; body condition scores, culling rate due to disease; cleanliness; cell counts, time budgets regarding lying, standing, feeding, respective behaviour rhythms, standing up behaviour; social interactions; avoidance distance towards humans [Waiblinger *et al* 2001]) we used estimated daily gain, fat score at slaughter and mortality as epidemiological welfare indicators to determine construct validity of the A-Index. Our approach is supported by Rushen (2003). He criticised assumptions behind experimental welfare studies since it is unlikely that the effects of housing type on animal welfare can be isolated from, and studied independently of, the effects due to nutrition, management, etc. He pointed out a need to use epidemiological methods in welfare studies.

Construct validity and sensitivity of the A-Index can be considered to be reasonable because estimated daily carcase

https://doi.org/10.1017/S0962728600000828 Published online by Cambridge University Press

gain increased, and the prevalence of high fat scores as well as mortality decreased with an increasing A-Index score, although the comparison criteria used are not sensitive to all aspects of welfare, eg fear, pain or frustration. Approximately 1 g difference in daily carcase gain and 0.1% difference in prevalence of fat scores or mortality corresponded with a single score difference in A-Index. In this sense the A-Index seemed able to reflect reasonably small changes in on-farm welfare. However, due to wide individual variation of measured outcomes, the A-Index cannot be used as a precise predictor for studied outcomes.

Conclusion and animal welfare implications

Summated scales are very popular tools to assess overall welfare and are easily understood by non-scientists. Partial scores can be used to point out strong and weak points of each farm assessed, and thus used for animal welfare advisory purposes. The overall score allows comparisons between animal units while an absolute judgement of a farm, independently of the others, can still be made (Botreau *et al* 2007a). Overall scores offer an opportunity to use a wide range of parameters enabling study of the multidimensional nature of animal welfare.

However, summated scales suffer from several limitations. They may allow compensation where compensation should be restricted and they do not favour situations of compromise (Botreau *et al* 2007a). A farm that obtains good or average scores for a summated scale may still have very low scores in individual items, and thus problems, regarding some animal welfare measures. To avoid problems associated with over compensation, minimum resource-based requirements were included in the A-Index following legislative or quality requirements in accordance with previous studies (Keeling 2005). Space allowance was the only minimum requirement significantly associated with the outcomes of the study. It can be argued that animal-based minimal requirements could be the most appropriate to avoid major welfare problems due to over-compensation.

Measures used to assess animal welfare are not direct measures of mental state but only indices that need to be interpreted in terms of welfare (Botreau *et al* 2007b). These indices are human constructs that are inherently laden with many of our values (Fraser 1995). In this respect, methods used widely in psychometrics and social sciences would represent a substantial advantage in choosing the most appropriate parameters, that is, items for overall assessment scales. The described method is a way to respond to claims, presented by Rushen (2003), questioning whether welfare is a measurable property of an animal. Welfare can be determined as a concept concerning a prolonged mental state, resulting from how the animal experiences its environment over time. Development and regular use of summated scales to measure overall animal welfare can be seen to be valuable because as long as separate behavioural, physiological and health parameters are measured, welfare statements are based on theoretical constructions or implicit opinions. Explicit welfare statements could be based on direct measurements of overall welfare by summated scales constructed and validated by methods widely used in psychometrics, and educational and social sciences.

Acknowledgements

The authors gratefully acknowledge V Tuovinen and P Aho for starting the welfare programme in Atria Ltd, J Alatalo and all personnel in Atria Ltd for organising the welfare programme as a part of quality work in Atria Ltd, S Finnilä, A Korhonen, K Mäki-Paavola and S Saastamoinen for on-farm scorings, Laura Hänninen and other members of the animal welfare research school at the University of Helsinki and personnel of the University of Kuopio for advice on welfare questions, personnel at MTT Agrifood Research Finland for consulting on meat production, personnel of the Finnish Food Safety Authority for advice concerning animal health, I Dohoo and H Stryhn from the University of Prince Edward Island and W. Martin from the University Guelph, Canada for giving good instructions for epidemiological analyses and J Metsämuuronen from the International Methelp ky for Test Theory consultation. The development process was funded by the European Agricultural Guarantee Fund (EAGF) and Atria Ltd. The scientific study was funded by Suomen Eläinlääketieteen Säätiö and Mercedes Zachariassenin Rahasto.

References

Amon T, Amon B, Ofner E and Boxberger J 2001 Precision of Assessment of Animal Welfare by the TGI 35 L, Austrian Needs Index. *Acta Agriculturae Scandinavica, Section A-Animal Sciences 51*: 114-117

Andersen HR, Jensen LR, Munksgaard L and Ingvartsen KL 1997 Influence of floor space allowance and access sites to feed trough on the production of calves and young bulls and on the carcass and meat quality of young bulls. *Acta Agriculturae Scandinavica Section A, Animal Science 47*: 48-56

Andersson M 1984 Drinking water supply to housed dairy cows. Influence on performance and behaviour of flow rate, water temperature, number of bowls, restriction in availability and social rank. *Swedish University of Agricultural Sciencies, Department of Animal Nutrition and Management. Uppsala. Report 130*: 54-57

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. *Livestock Production Science 61*: 179-192

Beyer S 1998 *Konstruktion und Überprüfung eines Bewertungskonzeptes für pferdehaltende Betriebe unter dem Aspekt der Tiergerechtheit.* Dissertation, Justus Liebig-Universität, Giessen, Germany. [Title translation: Construction and testing of a concept to measure horse farms from animal rights point of view]

Botreau R, Bonde M, Butterworth A, Perny P, Bracke MBM, Capdeville J and Veissier I 2007a Aggregation of measures to produce an overall assessment of animal welfare. Part 1: A review of existing methods. *Animal 1*: 1179-1187

Botreau R, Bracke MBM, Perny P, Butterworth A, Capdeville J, Reenen CG and Veissier I 2007b Aggregation of measures to produce an overall assessment of animal welfare. Part 2: Analysis of constraints. *Animal 1*: 1188-1197

Bracke MBM 2007 Animal-based parameters are no panacea for on-farm monitoring of animal welfare. *Animal Welfare 16*: 229-231

Bracke MBM, Spruijt BM and Metz, JHM 1999 Overall animal welfare assessment reviewed. Part 1: Is it possible? *Netherlands Journal of Agricultural Science 47*: 279-292

Broom DM 1991 Animal welfare: concepts and measurement. *Journal of Animal Science 69*: 4167-4175

Capdeville J and Veissier I 2001 A method of assessing welfare in loose-housed dairy cows at farm level, focusing on animal observations. *Acta Agriculturae Scandinavica, A, 51, Supplement 30*: 62-68

Cronbach LJ 1951 Coefficient alpha and the internal structure of tests. *Psychometrika 16*: 297-334

Dallman MF, Pecoraro N, Akana SF, la Fleur SE, Gomez F, Houshyar H, Bell ME, Bhatnagar S, Laugero KD and Manalo S 2003 Chronic stress and obesity: A new view of comfort food. *Proceedings of the National Academy of Sciences 100*: 11696-11701

Dawkins MS 1980 *Animal Suffering: The Science of Animal Welfare*. Chapman and Hall: UK

DeGrazia D 1996 *Taking Animals Seriously: Mental Life and Moral Status*. Cambridge University Press: Cambridge, UK

DeVellis RF 2003 *Scale Development: Theory and Applications*. Sage Publications: Thousand Oaks, USA

Dohoo I, Martin W and Stryhn H 2003 *Veterinary Epidemiologic Research*. AVC Incorporation: Charlottetown, Prince Edward Island, Canada

Duncan IJH 1996 Animal welfare defined in terms of feelings. *Acta Agriculturae Scandinavica 27*: 29-35

Fraser D 1995 Science, values and animal welfare: exploring the 'inextricable connection'. *Animal Welfare 4*: 103-117

Hannan J and Murphy P 1983 *Comparative Mortality and Morbidity Rates for Cattle on Slatted Floors and in Straw Yards*. Current Topics in Veterinary Medicine and Animal Science: The Hague, The Netherlands

Hassall S, Ward W and Murray R 1993 Effects of lameness on the behaviour of cows during the summer. *The Veterinary Record 132*: 578-580

Herva T, Huuskonen A, Virtala A, Saatkamp H and Peltoniemi OAT On-farm welfare and estimated daily carcass gain of slaughtered bulls. *Acta Agriculturae Scandinavica, Section A, Animal Science 59(2)*: 104-120

Ingvartsen KL and Andersen HR 1993 Space allowance and type of housing for growing cattle. A review of performance and possible relation to neuroendocrine function. *Acta Agriculturae Scandinavica, Section A, Animal Science 43*: 65-80

Jóhannesson T and Sørensen J 2000 Evaluation of welfare indicators for the social environment in cattle herds. *Animal Welfare 9*: 297-316

Keeling LJ 2005 Healthy and happy: animal welfare as an integral part of sustainable agriculture. AMBIO. *A Journal of the Human Environment 34*: 316-319

Lindström T and Redbo I 2000 Effect of feeding duration and rumen fill on behaviour in dairy cows. *Applied Animal Behaviour Science 70*: 83-97

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

Munsterhjelm C and Herva T 2003 Hyvinvointimittarit osaksi laatutyötä? *Finnish Veterinary Journal 91*: 79-81. [Title translation: Should welfare scores be included in the quality schemes?]

Munsterhjelm C, Valros A, Heinonen M, Halli O and Peltoniemi OAT 2006 Welfare index and reproductive performance in the sow. *Reproduction in Domestic Animals 41*: 494-500

Nunnally JC and Bernstein IH 1994 *Psychometric Theory*. McGraw-Hill: New York, USA

Phillips CJC, Morris ID, Lomas CA and Lockwood SJ 2000 The locomotion of dairy cows in passageways with different light intensities. *Animal Welfare 9*: 421-431

Ruis-Heutinck LFM, Smits MCJ, Smits AC and Heeres JJ 2000 Effects of floor type and floor area on behaviour and carpal joint lesions in beef bulls. Improving health and welfare in animal production. *Proceedings of Sessions of the EAAP Commission on Animal Management and Health* pp 29-36. 21-24 August 2000, The Hague, The Netherlands

Rushen J 2003 Changing concepts of farm animal welfare: bridging the gap between applied and basic research. *Applied Animal Behaviour Science 81*: 199-214

Schrader L, Hoth HR, Winterling C, Brodmann N, Langhans W, Geyer H and Graf B 2001 The occurrence of tail-tip alterations in fattening bulls kept under different husbandry conditions. *Animal Welfare 10*: 119-130

Scott EM, Nolan AM and Fitzpatrick JL 2001 Conceptual and Methodological Issues Related to Welfare Assessment: A Framework for Measurement. *Acta Agriculturae Scandinavica, Section A-Animal Sciences 51*: 5-10

Testa MA and Simonson DC 1996 Assessment of Quality-of-Life Outcomes. *New England Journal of Medicine 334*: 835-840

Waiblinger S, Knierim U and Winckler C 2001 The development of an epidemiologically based on-farm welfare assessment system for use with dairy cows. *Acta Agriculturae Scandinavica, Section A-Animal Sciences 51*: 73-77

Webster AJF, Smith JS and Brockway JM 1972 Effects of isolation, confinement and competition for feed on the energy exchanges of growing lambs. *Animal Production 15*: 189-201

Whay HR, Main DC, Green LE and Webster AJF 2003 Assessment of the welfare of dairy cattle using animal-based measurements: direct observations and investigation of farm records. *The Veterinary Record 153*: 197-202

Winckler C, Capdeville J, Gebresenbet G, Hörning B, Roiha U, Tosi M and Waiblinger S 2003 Selection of parameters for on-farm welfare-assessment protocols in cattle and buffalo. *Animal Welfare 12*: 619-624

https://doi.org/10.1017/S0962728600000828 Published online by Cambridge University Press