

EVALUATION OF WELFARE STATE BASED ON INTERPRETATION OF MULTIPLE INDICES

E M Scott^{*}, J L Fitzpatrick[†], A M Nolan[‡], J Reid[†] and M L Wiseman[†]

^{*} Department of Statistics, University of Glasgow, Glasgow G12 8QW, UK

[†] Institute of Comparative Studies (Department of Veterinary Clinical Studies), University of Glasgow, Glasgow G12 8QW, UK

[‡] Institute of Comparative Studies (Department of Veterinary Pre-Clinical Studies), University of Glasgow, Glasgow G12 8QW, UK

* Contact for correspondence and requests for reprints: marian@stats.gla.ac.uk

Abstract

Animal Welfare 2003, **12**: 457-468

Welfare is a multidimensional construct and its quantification is of major scientific, societal and economic importance in veterinary medicine. The construction of indices that measure welfare validly and reliably remains a considerable challenge. A general methodology for constructing welfare indices can be adapted from human medicine (in particular, from methodologies to assess Quality of Life [QoL]) and modified to reflect the fact that all assessments of animal welfare must be observer-based. The methodology is based on the creation of individual, composite indices for distinct dimensions/domains of welfare such as pain, disease, or stress. The domains include behavioural, physiological and biochemical markers. We have established QoL methodologies in the assessment of acute and chronic pain in dogs and generalised this approach to farm animal welfare. We describe the development of a questionnaire with seven behavioural categories which are used to create a single pain score to assess acute pain in dogs. For chronic pain in dogs, a structured questionnaire with over 100 items has been devised, which the owner completes by indicating degree of agreement with each item using a seven-point Likert scale. The welfare measure includes pain as an integral component as well as husbandry, behavioural and physiological/biological measures. In each case, a profile of the individual indices can be studied and compared over time or among observers. These indices may also be combined to form a single composite welfare measure, should this be appropriate, using scaling models. In the welfare setting, we have both causal and indicator variables — and indeed, for farm animals, the causal variables may be sufficient cause for poor welfare (eg the presence of disease or inadequate husbandry).

Keywords: *animal welfare, animal welfare assessment, health-related quality of life, psychometric theory, scaling models*

Introduction

“When you cannot measure it, when you cannot express it in numbers — you have scarcely in your thoughts, advanced to a stage of science, whatever the matter may be”, Lord Kelvin (1891).

In all facets of animal welfare, as in human health and quality of life (QoL), the unambiguous and valid measurement of key attributes is of critical importance. Welfare is a

complex construct and in this paper we present a methodological framework, using principles developed and applied in human-health-related quality of life (HRQoL) studies, for the development of animal welfare instruments. We indicate how psychometric and metrological principles can be applied in veterinary medicine and present case studies in acute and chronic pain in dogs and in dairy cow well-being.

Why is animal welfare important?

Welfare and QoL in animals is of considerable and growing importance from social, political, ethical and scientific viewpoints. Welfare is an abstract concept which is related to QoL, a concept that has been formally recognised and used in human medicine, where there has been considerable development of scales for its measurement (Bowling 1991; Skevington 1998). Quality of life, and specifically health-related quality of life, has been defined as “a state of complete physical, mental and social well-being, not merely the absence of disease” (WHO 1948). In the animal and veterinary sciences, similar views concerning pain, welfare and QoL are slowly emerging but are complicated and hampered by the fact that the ‘patient’ cannot communicate verbally and evaluation must, therefore, be based on signs, not symptoms, and the rating is proxy, carried out by an observer (veterinarian, owner, stockman or farmer). The growing concern and need for the well-being of animals to be duly recognised is highlighted in public concern for farm animal welfare and in the considerable recent growth of numerous farm assurance schemes (FAWC 2001). Disease is a major constraint to productivity and profitability of farm animal production in terms of adverse animal welfare (disease, abnormal behaviour and poor husbandry). Improved disease diagnosis and recognition of pain and adverse welfare in farm animals is essential to a progressive farming industry which takes into account the viewpoint of both the retailers and the consumers of products of animal origin.

How is welfare defined?

Welfare is a complex construct that combines both subjective and objective aspects of the conditions of life for animals. The Five Freedoms (FAWC 2001) are a well-established set of propositions underlying good farm animal welfare. They encompass freedom from hunger and thirst, freedom from discomfort, freedom from pain, injury and disease, freedom from fear and distress, and the freedom to express normal behaviour. Obvious aspects of animal welfare include animal health and disease, behaviour, pain, management practices, housing and husbandry. It is clear that in all facets of animal welfare, as in human health and QoL, the unambiguous and valid measurement of the key attributes is of critical importance. Assessment of welfare must, therefore, involve the assessment of a number of attributes or domains, many of which can only be assessed qualitatively.

We propose a general methodology for assessment and quantification of animal welfare intended to meet these requirements. The key steps in the methods proposed follow a pattern previously used in the development of instruments for the measurement of pain and QoL in humans (Melzack 1975; Bech 1999) and which are broadly based on psychometric principles.

The metrological principle which this paper proposes encompasses the creation of a single, composite indicator of welfare formed from sets of distinct, observable welfare components through the use of a scaling model. A variety of scaling models exists including direct or indirect estimation models, and we briefly discuss their merits. Further, the original welfare components can also be studied through analysis of the profile of measurement. The

distinct welfare components used vary depending on the species involved, but might include disease state, husbandry, and behaviour as more quantitative measures, and could also include more subjective expert judgements concerning the animals' demeanour or emotion.

What are the types of measurement and what properties do we want them to have?

The basic role of measurement is to assign numerical values to the attribute of interest or to classify an object on the basis of that attribute. Thus we can consider as our goal the quantification and categorisation of the welfare state of the animal in such a way that the derived welfare measure is valid and reliable, simple to use and responsive to change.

Units of measurement

Measurement can be considered at four different levels — nominal, ordinal, interval and ratio — and before creating a welfare scale, we must first consider the measurement level which is most appropriate and desirable. This decision should be based on the required sensitivity of the instrument, which is linked to the definition of welfare (and to the definition of what are significant differences) and the precision required for the actual measurement.

Nominal measurement occurs when the values assigned form categories which have no inherent ordering, for example sex. The categories serve merely as labels for the observations made and do not quantify them. Thus, this would be inappropriate for a welfare scale.

Ordinal measurement gives responses that are ordered and categorical in nature, such as poor, good, better, best, or mild, moderate, severe. This level of measurement provides no information on the differences between the categories, only on their relative ordering. Interval level measurement is created when the measurement is continuous rather than categorical in nature and the differences between a response and a constant are known. As well as providing the relative ordering, the difference between two points on the scale is known absolutely. Ratio measurement requires that an absolute zero of the attribute be defined, and thus is not appropriate for welfare measurement.

Ordinal and interval level measurements are both practicable and desirable for the assessment of farm animal welfare. Ordinal scales have been the most frequently created scales for welfare assessment, yet they present difficulties in analysis and interpretation. An ordinal scale is rather general but may offer the precision required, although it is important to be aware that its sensitivity and responsiveness to change are compromised if the ordered categories are broad. There must be a careful consideration of the number and definition of ordered levels. Interval level measurement is more demanding to create, but provides more precise measurement.

Validity and reliability

Any instrument or measurement tool must also be valid and reliable. If a scale is to be valid, it must measure the attribute for which it was developed. There are a number of different forms of validity (Streiner & Norman 1995; Scott *et al* 2001) some of which (face and content) are subjectively assessed from expert judgement. Other forms of validity (criterion and construct) can be formally assessed based on 'a gold standard measure', or some other independent measure. In the case of welfare, which cannot be measured directly, for which no gold standard exists and for which there are indeed different definitions, the validity must be investigated thoroughly before the instrument can be accepted for general use.

The reliability of an instrument quantifies the errors inherent in the scores generated and is based on a measurement model. The simplest and classic form of the model assumes that

there is a true underlying welfare score for each 'individual' and a measurement error associated with that score. The reliability is usually quantified as a dimensionless coefficient lying between 0 and 1 and is simply the ratio of the variability observed between the subjects' scores and the overall observed variability. As with validity, the scale developer must assess the reliability of the proposed instrument.

Classic measurement tools

Classic measurement tools widely used in veterinary medicine include the Simple Descriptive Scale (SDS), where the response is ordinal. The classic response set typically includes the categories 'none', 'mild', 'moderate' and 'severe'. The Numerical Rating Scale (NRS) is also widely used; this scale is also ordinal, although it is frequently treated as interval. The response is an integer from 0 to 9. Finally the Visual Analogue Scale (VAS) is widely used, and involves marking the level of the attribute on a line of fixed length, anchored at one end by 'none' and the other by 'level could not be higher'. These measurement instruments are uni-dimensional and so are likely to measure only intensity, whereas welfare — and pain — are likely to be multi-dimensional experiences, and so they cannot adequately capture its complexity (Williams *et al* 2000; Clark *et al* 2002).

These instruments are, however, simple to use and are generally considered valid and reliable (at least, when used in a self-rating mode in human medicine). However, their use in veterinary medicine by observers has often shown significant inter-observer variability (Holton *et al* 1997). With increasing emphasis on the use of objective measurements to assure good husbandry, management systems and welfare on farms, and recognising that pain is a key contributor to poor welfare, it was considered timely to probe the perceptions of those working closely with sheep and cattle in relation to common inflammatory diseases, routine elective procedures and associated pain.

Perceptions of pain in sheep and dairy cattle

For cattle, inflammatory disease is probably the major cause of pain. The most significant inflammatory diseases likely to be associated with pain in cattle are mastitis (Kossaibati *et al* 1998) and inflammatory limb lesions, including sole ulceration, white line disease and acute interdigital tissue infection, leading to lameness (Whay *et al* 1998). Both mastitis and lameness are diseases of major importance in the UK with annual incidences of 40% and 54%, respectively (Clarkson *et al* 1996; Kossaibati *et al* 1998). Studies have shown that it is possible to scale disease severity for mastitis (Fitzpatrick *et al* 1999) and for lameness (Whay *et al* 1998), although no pain scales have been currently described in cattle for these or other common diseases. Other causes of pain include routine elective procedures such as castration or dehorning/disbudding of calves (Moloney *et al* 1995; Kent *et al* 1996). There are limited data (Watts 2000) on the perceptions amongst veterinary surgeons of pain intensity associated with various diseases and elective procedures in cattle.

Many factors contribute to poor welfare in sheep, including husbandry, poor nutrition and disease. Diseases likely to be most important in this respect include the inflammatory diseases, many of which are associated with pain. Inflammatory diseases that cause lameness, including footrot, and skin lesions, including sheep scab, are easily identified as causes of pain; however, diseases such as mastitis, which are less obvious to an observer, may also cause pain, and are responsible for increased premature culling and thus contribute to poor welfare. Pain associated with castration and tail-docking in lambs, and the extent of the alteration in specific behaviours associated with different methods of tail-docking and

castration, have been well described (Moloney *et al* 1993; Moloney & Kent 1997; Thornton & Waterman-Pearson 1997; Kent *et al* 2000; Price & Nolan 2001). Furthermore, the pathophysiological alterations in pain processing associated with inflammatory disease in humans have also been recorded in sheep with inflammatory disease (Ley *et al* 1995; Welsh & Nolan 1995; Kent *et al* 1998; Dolan & Nolan 2000).

Sheep husbandry practices may contribute to the difficulty in identifying signs of pain associated with inflammatory disease, since the animals are seldom observed individually. Consequently we designed a survey to study observers' perceptions of pain in both sheep and cattle.

Methods

A group of participants at the Sheep Veterinary Society meeting (Oxford 2001) and a group of predominantly veterinary surgeons with expertise in cattle practice attending a British Cattle Veterinary Association meeting (Glasgow 2000) were asked about their perceptions of pain for a number of common procedures and diseases. Their responses were recorded using an interactive scoring system. Participants were also asked their age, sex, time qualified and area of veterinary experience.

The participants were asked to score the intensity of pain associated with the common procedures and conditions detailed below using an eleven-point numerical rating scale (NRS), from 0 to 10, where 0 = no pain and 10 represents the worst pain imaginable.

For sheep, the conditions and procedures were:

1. castration of lambs by a) rubber ring or b) open surgery
2. tail-docking of lambs
3. Caesarean section
4. 'footrot'
5. 'flystrike'
6. chronic mastitis

For cattle, the conditions and procedures were:

1. castration of calves by a) rubber ring or b) bloodless castrators or c) open surgery
2. Caesarean section
3. 'foul-in-the-foot'
4. solar ulcer
5. mastitis

Results for sheep

The pain scores (and their distributions) associated with the procedures and diseases described above are shown in Table 1. The vast majority (96–100%) of respondents considered that some pain was associated with all of these procedures and conditions.

The data indicate a ranking of pain intensity associated with the various procedures; for example, castration by rubber ring (median pain score 6) was perceived to be more painful than castration by open surgery (median pain score 5), which was perceived to be more painful than tail-docking (median pain score 4) (Table 1). The median pain score for Caesarean section was 4, with the inter-quartile range (IQR) being 3–6. The pain scores for 'footrot' were moderate to high with a median score of 6 and an IQR of 5–7.5. The

distribution of pain scores for 'flystrike' was more uniform, between 3 and 6 with a median value of 5, and for chronic mastitis (median value 4) there was an uneven spread with low to moderate scores being more common than for 'footrot' or 'flystrike'. Thus, of the diseases considered, 'footrot' was reported to be the most painful, while chronic mastitis was thought to be the least painful by respondents.

Table 1 Pain intensity summary statistics for sheep. Respondents (n = 77) scored the pain associated with each 'condition' using an eleven-point numerical rating scale (0 = no pain; 10 = maximum pain).

Condition	Mode	Median score	Inter-quartile range
<i>Castration by rubber ring</i>	7 or 8	6	5–8
<i>Castration by open surgery</i>	5	5	4–7
<i>Tail docking by rubber ring</i>	3	4	3–6
<i>Caesarean section</i>	3	4	3–6
'Footrot'	7	6	5–7.5
'Flystrike'	5	5	3–6
<i>Chronic mastitis</i>	3	4	3–6

Results for cattle

For all the cattle procedures and conditions, the audience considered that some pain was associated with them (Table 2). The pain scores for castration using the three different methods indicated that there was considerable variation in opinion but that, overall, most people considered the pain level to be moderate (scores between 4 and 7). A higher proportion of high pain scores was reported for castration by the rubber ring method and the bloodless castrator than for the open surgery method. The majority of the pain scores for Caesarean section were reported in the range 4–8. Pain scores for both conditions causing lameness (ie 'foul-in-the-foot' and solar ulcer) were considered to be rather painful, with the majority of respondents scoring between 6 and 8. It is interesting to observe that very similar patterns of scores were allocated to both lameness conditions. Mastitis scores showed a wider range than all other conditions.

Table 2 Pain intensity summary statistics for cattle. Respondents (n = 79) scored the pain associated with each 'condition' using an eleven-point numerical rating scale (0 = no pain; 10 = maximum pain).

Condition	Mean	Median score	Inter-quartile range
<i>Castration by rubber ring</i>	5.7	6	4–7
<i>Castration by bloodless castrators</i>	6.0	6	5–7
<i>Castration by open surgery</i>	5.2	5	4–6
<i>Caesarean section</i>	5.6	6	4–7
'Foul-in-the-foot'	6.5	7	6–8
<i>Solar ulcer</i>	6.8	7	6–8
<i>Mastitis</i>	4.9	5	4–6

In summary, there was a general agreement that routine procedures and conditions in sheep and cattle were associated with pain and that there was a perceived hierarchy of pain with these conditions. Considerable observer variation among the perceived intensities of pain associated with the different conditions was recorded. Such variation is not unexpected and is in common with many other subjective and qualitative judgements.

Health-related QoL and its application to pain and welfare measurement

The psychometric methods established by psychologists and psychiatrists to measure multiple-attribute 'constructs' using formally assessed, structured questionnaires, or instruments, have increasingly been adopted in the measurement of human pain (McArthur *et al* 1989; Jensen & Karoly 1992), particularly chronic pain (Vlaeyen *et al* 1990; Landgraf & Abetz 1996) and QoL (Schipper 1990; Naughton *et al* 1996). Our objective is therefore the modification of such an approach for use in pain and welfare in the veterinary setting.

Quality of life is considered to be a multi-dimensional construct focussing on, among other things, physical functioning, social well-being and general health. Items to be included are frequently identified by experts through survey and consultation. Instruments devised to measure QoL typically take the form of a structured questionnaire and include many items (questions) linked to each of the domains of interest. Often each item has a simple ordered response: not at all; a little; quite a bit; very much. Items may be weighted to represent their importance, and scores are created by summing the individual items (using equal or unequal weights). Individual items may themselves comprise multi-dimensional attributes such as health or pain which will require sub-scales to be constructed independently. All the items must be defined in detail, since the observation and recording protocol is a vital part of the instrument development when used by observers as proxy raters.

Presenting the results

The results from the structured questionnaire can be presented in a number of ways. First, a composite score (from combining the individual item scores) can be formed using a scaling model, while in the second option, scores for the individual dimensions are formed and not combined. Thus, in this latter case, a profile score is formed and the domains can be interpreted in the context of the others.

Scaling models

A scaling model is a technique that allows weights to be devised for the items included in a scale reflecting the level of welfare associated with the given item. There are two main types of scaling models: direct or subjective estimation techniques; and indirect or discriminant techniques (Nunnally & Bernstein 1994). The direct or subjective estimation techniques are based on the developers' best subjective estimate of the weights that should be assigned to the items. The true relative importance can be explored by considering all the pairs of items but this is seldom done, hence the validity of the weighting scheme may be in question, and consequently this technique would not be suitable for a composite welfare index.

In indirect or discriminant techniques, the weight for each item is derived from experimental observations. Two of the most commonly used indirect scaling models, the equally weighted and paired comparison models, are discussed below.

The equally weighted model is the simplest of all scaling models. This model assumes an equal weight for each of the items included in the measurement instrument and assigns a score of 1 to each item. The total combined score represents the number of items observed when the assessment is made.

The method of paired comparisons was derived from the classical law of comparative judgement proposed by Thurstone (1927). This scaling model assumes that the items included in an instrument are correlated with an intensity of the attribute of interest (eg welfare) and that the intensity associated with each item follows a normal distribution. Hence, the best estimate of a weight for any item is its associated mean welfare intensity.

Since the attribute of interest (eg welfare) cannot be measured directly, the intensity associated with each item can only be judged relative to the other items within each domain.

The weights for each item included in the scale are calculated using a panel of judges. Each judge compares each pair of items included in the domains and identifies which item is associated with the highest degree of the attribute. These comparisons allow the items to be ordered relative to each other and weights for each item can be estimated by transforming the observed proportions (Streiner & Norman 1995). The total score for any instrument can then be calculated by adding together scores for the items observed.

Applications

Acute clinical pain in dogs (Holton *et al* 1997, 1998, 2001)

The approach taken in the development of the acute pain instrument followed that described by Melzack (1975), but was modified appropriately to take account of the additional complications presented by the lack of language of the species.

A bank of words and expressions describing behaviours which were thought to be signs of pain in dogs was collected from a total of 69 practising veterinary surgeons. Each veterinary surgeon was asked to list all behavioural and physiological signs he or she thought would be shown by a dog suffering pain of any origin. A total of 279 words and expressions were collected. An initial reduction phase resulted in a final list of 47 expressions consisting of 39 behavioural and eight physiological signs of pain. The procedure of identifying lists of items through survey of a large number of experts should, in this way, ensure that the scale has face and content validity.

After reduction, a possible categorisation of the words and expressions was explored by a focus group of veterinary surgeons. This resulted in a proposed grouping into nine domains. The categories were: demeanour; response to people; response to food; posture; mobility; activity; response to touch; attention to painful area; and vocalisation. There was general agreement about the allocation of the expression to categories.

A scaling model was then chosen that would allow pain to be quantified and provide measurement on an interval level. The most appropriate method was considered to be the method of paired comparisons (or matched pairs) derived from the classical law of comparative judgement proposed by Thurstone (1927). This scaling model assumes 1) that the scale items (in this case, behaviours) are broadly correlated with an intensity or strength of the attribute of interest (ie pain intensity); and 2) that each item follows an underlying normal distribution which reflects the variation in the intensity of the attribute that is associated with that item. Consequently, the best estimate of a weight for any item is the mean intensity associated with that item. The process of estimating the weights involved gathering information on the perceived intensity of the items relative to each other, then estimating the mean intensity. A Thurstone matched pairs model was used to define weights for the items, to create a composite score from 0 to 10. This prototype acute pain instrument has undergone a series of validation and reliability-testing experiments and a short-form version is under construction (Holton *et al* in preparation).

Creation of a chronic pain QoL instrument in dogs (Wiseman *et al* 2001)

A preliminary study, consisting of informal interviews with owners and with veterinary surgeons, had provided some evidence that chronic pain in dogs was associated with a wide range of behavioural disturbances and that these changes could be observed and reported by

owners (Wiseman *et al* 2001). The first stage in the development of an instrument for chronic pain therefore established the domains of behaviour relevant to HRQoL measurement in dogs, and collected items to describe these.

The information was gathered directly from owners of dogs considered to be suffering chronic and painful conditions. Disturbances in a total of 32 types of behaviour were reported, including changes in activity, mobility, agility, daytime sleeping/resting, attitude and demeanour, stamina, playfulness, pain-related vocalising, facial expression, sociability toward family members, keenness to exercise, drinking and posture, as well as inconsistency in behaviour. Qualitative interpretation of the data was also undertaken and yielded four findings. All owners had confidence in their awareness of their dogs' behaviour, they compared behaviour with their dogs' 'normal' state, they were capable of remarking gradations in behaviour, and they interpreted some changes in behaviour as indicators of internal mood state.

The behaviour changes were grouped, according to apparent associations, into 11 dog HRQoL domains. The domains thus created were: activity, comfort, appetite, extroversion/introversion, aggression, anxiety, alertness, dependence, contentment, consistency, and agitation. Potential items were obtained by means of a series of descriptor-generating surveys involving over 200 dog owners visiting a small animal hospital. Each of these potential items was allocated to one of the HRQoL domains. A matrix of behavioural domains affected by chronic pain, and associated positive and negative descriptors, was thus created.

The prototype instrument asked owners to rate, on a seven-point scale from 0 to 6, how well each of 109 descriptors described their dog. A score of 6 would indicate 'the best' or 'the worst', depending on whether the descriptor was positive (generally associated with a pain-free state) or negative (generally associated with chronic pain).

The matrix of terms and domains created was then subjected to validation by an expert group, and the validation is being further explored during field testing of the prototype. A similar approach has been used in the development of human pain and HRQoL instruments (Melzack 1975; Juniper *et al* 1997; Armstrong *et al* 1999).

Although established processes for the construction of HRQoL instruments were used (Coste *et al* 1995; Streiner & Norman 1995; Juniper *et al* 1996; Johnston 1998; Guyatt 1999), compared to existing instruments for assessing human HRQoL by proxy, the resulting prototype instrument represents a novel approach which provides the proxy rater with an opportunity to provide detailed information about subtle changes in a simple way. It is currently undergoing tests in the University of Glasgow small animal clinics.

Assessing well-being in farm animals

A current BBSRC-funded project has as its objective the development of a well-being instrument for dairy cows. It includes domains covering disease, pain, behaviour, housing and husbandry. The actual items and proposed measurements are described in Table 3 and include a series of structured questionnaires for assessing husbandry, assessing behaviour (in the parlour, during housing and at pasture), and clinical assessment of animals. Individual cow records including milk production, somatic cell counts and treatments administered have been developed. The approach adopted here has followed psychometric principles, with identification by experts of clinical signs and behaviours associated with painful disease. As

well as these key components of animal health, disease and behaviour, husbandry has also been included. In the welfare setting, both causal and indicator variables are included and indeed the causal variables may be sufficient cause for poor welfare (eg the presence of disease or inadequate husbandry).

Table 3 Domains of dairy cattle wellbeing.

Health	Objective indicators
• Milk production	<ul style="list-style-type: none"> • acute phase proteins • hyperalgesia
Housing and husbandry	Disease (incidence, duration and intensity)
	<ul style="list-style-type: none"> • mastitis • clinical, sub-clinical • lameness • locomotion scoring • lesion identification
Behaviour	
• Lying, rising	
• Interactions with	
• other cows	
• stockpeople	

Discussion and animal welfare implications

We believe that the construction of a welfare instrument based on the principles evolved in QoL studies is practicable. Multi-dimensional and composite measurement scales have been shown to possess greater overall reliability and validity than subjective methods (Wright & Feinstein 1992; Nunnally & Bernstein 1994). The processes described above would ensure that the composite welfare measure (and its individual domains) would be valid (content, face, criterion and construct) and reliable. We believe that the approach described provides a firm and novel basis for the construction of either a composite, or a profile, of welfare indices that is urgently required for the modern veterinary practice.

The importance of the properties of the resulting tool and how it will be used cannot be over-emphasised and we believe that the use of metrological principles is widely applicable and allows the creation of a valid and reliable instrument for measurement of animal welfare.

Acknowledgements

The authors would like to thank the organisers of the British Cattle Veterinary Association and the Sheep Veterinary Society meetings for allowing the interactive sessions to be presented and the audiences for their participation. The financial support of Pfizer Ltd is acknowledged gratefully. L Wiseman was supported by the Roland Miller Postgraduate Fund. The BBSRC project is joint work with Professor Christine Nicol of University of Bristol and Ms Lesley Penman.

References

- Armstrong F D, Toledano S R, Miloslavich K, Lackman-Zeman L, Levy J D, Gay C L, Schuman W B and Fishkin P E 1999 The Miami pediatric quality of life questionnaire: parent scale. *International Journal of Cancer* 83(S12): 11-17

- Bech P** 1999 Health-related quality of life measurements in the assessment of pain clinic results. *Acta Anaesthesiologica Scandinavica* 43(9): 893-896
- Bowling A** 1991 *Measuring Health: A Review of Quality of Life Measurement Scales*. Open University Press: Open University, UK
- Clark W C, Yang J C, Tsui S-L and Clark S B** 2002 Unidimensional pain rating scales: a multidimensional affect and pain survey (MAPS) analysis of what they really measure. *Pain* 98: 241-247
- Clarkson M J, Downham D Y, Faull W B, Hughes J W, Manson F J, Merritt J B, Murray R D, Russell W B, Sutherst J E and Ward W R** 1996 Incidence and prevalence of lameness in dairy cattle. *Veterinary Record* 138: 563-567
- Coste J, Fermanian J and Venot A** 1995 Methodological and statistical problems in the construction of composite measurement scales: a survey of six medical and epidemiological journals. *Statistics in Medicine* 14: 331-345
- Dolan S and Nolan A M** 2000 Behavioural evidence supporting a differential role for group I and group II metabotropic glutamate receptors in spinal nociceptive transmission. *Neuropharmacology* 39: 1132-1138
- Farm Animal Welfare Council** 2001 Interim report on the animal welfare implications of Farm Assurance Schemes. FAWC: London, UK
- Fitzpatrick J L, Young F J, Eckersall P D, Logue D N, Knight C H and Nolan A** 1999 Mastitis — a painful problem. *Cattle Practice* 7(2): 225-226
- Guyatt G H** 1999 Measuring health-related quality of life in childhood: lessons from the workshop (discussion). *International Journal of Cancer* 83(S12): 143-146
- Holton L, Scott E M, Nolan A and Reid J** 1997 Comparison of three methods of pain measurement when used to assess clinical pain in dogs. *Journal of the American Veterinary Medical Association* 212: 61-66
- Holton L, Scott E M, Nolan A, Welsh E, Reid J and Pawson P** 1998 Relationship between physiological factors and clinical pain. *Journal of Small Animal Practice* 39: 469-474
- Holton L, Reid J, Scott E M and Nolan A** 2001 Development of a behaviour based scale to measure acute pain in dogs. *Veterinary Record* 148: 525-531
- Jensen M P and Karoly P** 1992 Self-report scales and procedures for assessing pain in adults. In: Turk D C and Melzack R (eds) *Handbook of Pain Assessment*. The Guildford Press: New York, USA
- Johnston C C** 1998 Psychometric issues in the measurement of pain. In: Finley G A and McGrath P J (eds) *Measurement of Pain in Infants and Children: Progress in Pain Research and Management, Vol 10* pp 5-20. IASP Press: Seattle, USA
- Juniper E F, Guyatt G H and Jaeschke R** 1996 How to develop and validate a new health-related quality of life instrument. In: Spilker B (ed) *Quality of Life and Pharmacoeconomics, Edn 2* pp 49-56. Lippincott-Raven: Philadelphia, USA
- Juniper E F, Guyatt G H, Streiner D L and King D R** 1997 Clinical impact versus factor analysis for quality of life questionnaire construction. *Journal of Clinical Epidemiology* 50: 233-238
- Kent J E, Jackson R E, Molony V and Hosie B D** 2000 Effects of acute pain reduction methods on the chronic inflammatory lesions and behaviour of lambs castrated and tail docked with rubber rings at less than two days of age. *Veterinary Journal* 160(1): 33-41
- Kent J E, Molony V and Graham M J** 1998 Comparison of methods for the reduction of acute pain produced by rubber ring castration or tail docking of week-old lambs. *Veterinary Journal* 155(1): 39-51
- Kent J E, Thrusfield M V, Robertson I S and Molony V** 1996 Castration of calves: a study of methods used by farmers in the United Kingdom. *Veterinary Record* 138: 384-387
- Kossaibati M A, Hovi M and Esslemont R J** 1998 Incidence of clinical mastitis in dairy herds in England. *Veterinary Record* 143: 649-653
- Landgraf J M and Abetz L N** 1996 Measuring health outcomes in pediatric populations: issues in psychometrics and application. In: Spilker B (ed) *Quality of Life and Pharmacoeconomics in Clinical Trials, Edn 2* pp 793-802. Lippincott-Raven: Philadelphia, USA
- Ley S J, Livingston A and Waterman A E** 1995 A field study of the effect of lameness on mechanical nociceptive thresholds in sheep. *Veterinary Record* 137: 85-87

- McArthur D L, Cohen M J and Schandler S L** 1989 A philosophy for measurement of pain. In: Chapman C R and Loeser J D (eds) *Advances in Pain Research and Therapy, Vol 12: Issues in Pain Measurement*. Raven Press: New York, USA
- Melzack R** 1975 The McGill Pain Questionnaire: major properties and scoring methods. *Pain* 1: 277-299
- Moloney V and Kent J E** 1997 Assessment of acute pain in farm animals using behavioral and physiological measurements. *Journal of Animal Science* 75(1): 266-272
- Moloney V, Kent J E and Robertson I S** 1993 Behavioural responses of lambs of three ages in the first three hours after three methods of castration and tail docking. *Research in Veterinary Science* 55(2): 236-245
- Moloney V, Kent J E and Robertson I S** 1995 Assessment of acute and chronic pain after different methods of castration of calves. *Applied Animal Behaviour Science* 46(1-2): 33-48
- Naughton M J, Shumaker S A, Anderson R T and Czajkowski S M** 1996 Psychological aspects of health-related quality of life measurement: tests and scales. In: Spilker B (ed) *Quality of Life and Pharmacoeconomics in Clinical Trials, Edn 2*. Lippincott-Raven: Philadelphia, USA
- Nunnally J C and Bernstein I H** 1994 *Psychometric Theory*. McGraw Hill Series in Psychology: New York, USA
- Price J and Nolan A M** 2001 Analgesia of newborn lambs before castration and tail docking with rubber rings. *Veterinary Record* 149(11): 321-324
- Schipper H** 1990 Quality of life: meaning and measurement. In: Foley K M (ed) *Advances in Pain Research and Therapy*. Raven Press: New York, USA
- Scott E M, Nolan A M and Fitzpatrick J L** 2001 Conceptual and methodological issues related to welfare assessment: a framework for measurement. *Acta Agriculturae Scandinavica (Section A — Animal Science)* 30: 5-11
- Skevington S M** 1998 Investigating the relationship between pain and discomfort and quality of life, using the WHOQoL. *Pain* 76: 395-406
- Streiner D L and Norman G R** 1995 *Health Measurement Scales: A Practical Guide to their Development and Use*. Oxford Medical Publications: Oxford, UK
- Thornton P D and Waterman-Pearson A E** 1997 Castration in young lambs produces changes in mechanical nociceptive threshold responses and behaviour as assessed by a dynamic and interactive visual analogue scale. *Journal of Veterinary Anaesthesia* 24: 1-41
- Thurstone L L** 1927 A law of comparative judgement. *Psychological Reviews* 76: 31-48
- Vlaeyen J W S, Pernet D F M, Kole-Snijders A M J, Schuerman J A, Van Eek H and Groenman N H** 1990 Assessment of the components of observed chronic pain behaviour: the Checklist for Interpersonal Pain Behaviour (CHIP). *Pain* 43: 337-347
- Watts S A** 2000 A survey of bovine practitioners attitudes to pain and analgesia in cattle. *Cattle Practice* 8: 361-362
- Welsh E M and Nolan A M** 1995 Effect of flunixin meglumine on the thresholds to mechanical stimulation in healthy and lame sheep. *Research in Veterinary Science* 58: 61-66
- Whay H R, Waterman A E, Webster A J F and O'Brien J K** 1998 The influence of lesion type on the duration of hyperalgesia associated with hindlimb lameness in dairy cattle. *Veterinary Journal* 156(1): 23-29
- Williams A C, Davies H T O and Chadury Y** 2000 Simple pain rating scales hide complex idiosyncratic meanings. *Pain* 85: 457-463
- Wiseman M L, Nolan A M, Reid J and Scott E M** 2001 Preliminary study on owner-reported behaviour changes associated with chronic pain in dogs. *Veterinary Record* 149: 423-424
- World Health Organisation** 1948 Constitution of the World Health Organisation. WHO Basic Documents: Geneva, Switzerland
- Wright J G and Feinstein A R** 1992 A comparative contrast of clinimetric and psychometric methods for constructing indexes and rating scales. *Clinical Epidemiology* 45: 1201-1218