PAIN MANAGEMENT IN ANIMALS – Eds Paul Flecknell and Jo Murrell

Chapter 6 Pain Assessment

Section a. Use of instruments to measure pain – general

M. Lesley Wiseman-Orr, E. Marian Scott, Jacky Reid and Andrea M Nolan

The need for pain measurement instruments

The measurement of pain can facilitate decision making regarding the treatment of individual animals, whether active or palliative. Such measurement can also be used in clinical trials, to judge the effectiveness of one treatment compared with another, or with none. Our ability to measure pain in a valid and reliable manner is essential then to meet the growing demand for evidence-based veterinary medicine and to recognise, treat and manage pain more effectively in animals. Clinical decision making also represents an important welfare issue, particularly as treatment options increase for companion animals, some associated with short- or long-term negative impacts. Greater choice of treatment options and increased affordability require demanding ethical decision-making in veterinary practice. An instrument that can be used with confidence to monitor and record pain status in an individual over time, and to provide data that will facilitate the selection and development of treatments with known effectiveness and impact, will benefit patient, client and veterinary practitioner.

The role of measurement is to assign numerical values to the attribute of interest or to classify an object on the basis of that attribute – to quantify or categorise an animal's pain in such a way that we can have confidence in the derived measure. William Thomson, Lord Kelvin of Glasgow, famously said that 'when you cannot measure it, when you cannot express it in numbers… you have scarcely, in your thoughts, advanced to the stage of science, whatever the matter may be' (Thomson, 1889). This hints at the rigorous thinking that the development of any instrument demands. In veterinary medicine, many behaviour-based pain scales have been constructed on an *ad hoc* basis but there is growing support for rigorous methods to be applied to the development and testing of pain measures for use in veterinary medicine in order to deliver valid and reliable pain measurement tools.

The measurement of pain

Pain is subjective, defined for both animals and people as a 'sensory and emotional experience' (Molony, 1997; Loeser and Treede, 2008). As the suffering associated with pain is a key motivator for its measurement, the goal of such measurement is primarily to measure pain's affective component. This is not a simple task. The complexity of the pain experience is even greater when the pain becomes chronic. Because chronic pain in people interacts in a complex way with a patient's social, psychological and physical well-being, many of the instruments now used to measure human chronic pain are concerned primarily with the patient's health-related quality of life (HRQL) and attempts have been made to develop similar instruments for animals (e.g. Heilm- Bjorkman et al, 2003; Yazbek and Fantoni 2005; Wiseman-Orr et al, 2004, 2006; Brown et al, 2007). HRQL instruments are designed to assess chronic pain's wide-ranging impacts, and also treatment effects and side-effects. They are increasingly valued as outcome measures in human medicine and are likely to become so in veterinary medicine as suitable instruments become available. It has been suggested that a more comprehensive understanding of animal pain, and in particular the affective component, may be of critical importance in the development of therapies for chronic and neuropathic pain, for animals and for people (Flecknell, 2008).

Instruments to measure pain and HRQL can be used to measure differences between patients at a point in time (discriminative purposes) or differences within a patient over time (evaluative purposes). They can be specific, focusing on particular conditions or populations, or they can be generic, designed to be used in a variety of contexts. Specific instruments may be more responsive to clinical change, but generic instruments can be valuable indicators of a range of impacts associated with disease and its treatment, and may be the only option when a patient is suffering from more than one condition.

In recent years there has been an exponential increase in the availability of instruments to measure pain in people. Instrument development has been firmly based upon the concept of pain as a complex, abstract and subjective construct. Psychometric methods are used to develop suitable structured questionnaires with formal scoring methodology, which were originally established by psychologists and psychiatrists to measure abstract, multiple-attribute constructs such as intelligence and personality. The gold standard measure of human pain is the self report. For those who are incapable of self-report, such as infants or cognitively impaired adults, instruments are designed for completion by a proxy or observer.

Instrument development

Psychometric methodologies for the development and testing of instruments are very well established (e.g. Streiner and Norman, 2008; Abell et al, 2009) and can be summarised as follows:

- specify measurement goals (and hence the ideal measurement scale) – what is to be measured, in which population and for what purpose;

- develop a pool of potential items for the instrument;
- select suitable items from the pool and subject to expert validation;
- incorporate selected items into an instrument, with consideration given to layout, response options, instructions to respondent and administration;
- pre-test the instrument to ensure that the target respondent can use the instrument correctly;
- field test the instrument, in order to evaluate its psychometric properties.

Instrument development is an iterative process, in which instruments are refined and re-tested with new populations in new contexts and for new purposes. The development of instruments to measure pain is a time-consuming undertaking, but the important contribution of the psychometric approach to such instrument development is widely recognised (Cook et al, 2003). By adopting a rigorous methodological approach to constructing pain and HRQL instruments which assures their validity (see below), veterinary practitioners can be more confident of managing and treating pain of all origins in animals under their care.

Instrument properties

Validity

The most important property of an instrument is its validity – the extent to which it measures what it is intended to measure – and consequently the first consideration when appraising potential pain measures should be ‘does this really measure pain’? In making a judgement about an instrument’s validity, it is helpful to consider the various kinds of validity that an instrument can possess. Instrument developers should seek evidence for validity of three principal kinds: criterion validity, content validity (face validity – the extent to which items appear ‘on the face of it’ to be measuring what the instrument is intended to measure – which is related to content validity, may or may not be sought) and construct validity (Streiner, 1993; Jensen, 2003).

Criterion validity is the agreement of a new instrument (or parts of it) with some existing ‘gold standard’. When no suitable gold standard exists, evidence is normally provided for content and construct validity.

The content validity of an instrument is the extent to which the attribute(s) of interest are sampled comprehensively by the instrument’s items, and the appropriateness of each of the items to the measurement of interest. This is largely established through the methodology used to collect and choose the items to be included in an instrument, but is often formally assessed by an independent group of ‘experts’ (Streiner and Norman, 2008). The items selected for instruments to measure pain and HRQL are frequently of two kinds: items that seek to capture variables which are likely to impact upon pain or HRQL (‘causal’ variables such as symptoms), and variables which manifest changes in those constructs (‘indicator’ variables that reflect how the subject feels). The latter may require the observer to interpret the animal’s behaviour qualitatively (e.g. the animal is ‘alert’ or ‘depressed’). Justification has been provided for the use of qualitative interpretation of animal behaviour as a means of obtaining information about its mental state (Wemelsfelder, 1997; Wemelsfelder *et al.*, 2001) and pain and HRQL instruments for use with animals can make use of this capability (e.g. Wiseman-Orr et al, 2004, 2006).

Respondent bias is a risk to valid measurement of pain, particularly when using an instrument with high face validity. For example, if scoring a dog with a cruciate repair shortly after surgery, a veterinary practitioner who has him/herself undergone similar surgery may be inclined to score the pain more severely than one who has not. This is compounded by the complexities of the relationship between an owner or carer and an animal, and those developing measures for use by parents and carers as proxy respondents in paediatric medicine face similar problems. Careful consideration should be given to this risk during selection from among existing instruments, or during development of new ones.

The construct validity of an instrument can be assessed in a variety of ways. Factorial validity is one kind of construct validity that requires the statistical analysis of correlations between responses to the items of an instrument. If an underlying factor structure fits the construct upon which the instrument was developed, which, in the case of pain and HRQL is not a readily observable physical but rather a hypothetical construct, then some evidence has been provided for the validity of the instrument and also for that hypothetical construct (Johnston, 1998). Evidence for the construct validity of an instrument is also provided when the scores obtained with the instrument fit the hypothetical construct upon which the instrument was developed by the extent to which the scores for different known groups or within groups over time can be predicted by that construct (Guyatt, 1993; Streiner, 1993; Johnston, 1998). For example, by showing that pain scores rise and fall predictably over time following surgery (Morton et al, 2005).

The importance of instrument validity cannot be overemphasised. However, an instrument cannot be said to ‘be valid’: it can only be shown to have validity for particular purposes, with defined populations and in specified contexts (Streiner and Norman 2008). Therefore, a key question to ask during instrument selection is ‘what evidence is there that this instrument can measure what I want to measure in the population that I want to measure and for my particular purpose? Other key considerations for instrument selection are listed at the end of this section.

Type of measurement scale

The measurement from an instrument can have nominal, ordinal, interval or ratio scale properties depending on the design of the instrument. Ordinal measurement, which provides information about how individuals relate to one another in relative terms, and interval measurement, with which individuals can be placed on a scale of equal units, are practicable and desirable for the assessment of pain. Ordinal scales have been the most frequently created and may offer enough precision in some circumstances, although their sensitivity and responsiveness to change is compromised if the ordered categories are broad. Examples include the simple descriptive scale (SDS) with categories of None, Mild, Moderate and Severe. Interval level measurement is more demanding to create (e.g. Morton et al, 2005), but provides more precise and meaningful measurement, and is likely to have increased sensitivity and responsiveness.

Response options

Each item in a questionnaire-like instrument is accompanied by an answer option or options. These may be dichotomous (yes/no), categorical (e.g. mild/moderate/severe), ordinal (e.g. numerical Likert-type scale) or even more complex. If responses are likely to lie on a continuum, it is important that respondents have the opportunity to answer in this way to ensure minimum loss of information and to minimise error (Streiner and Norman, 2008). For example, for a question about anxiety, more and better information will be captured by offering the opportunity to indicate an amount of anxiety than would be captured by simply asking if the subject is or is not anxious.

Different types of scale are commonly used for the direct estimation of assumed continuous variables, including numerical rating scales (NRS), visual analogue scales (VAS), adjectival scales (with or without a VAS) and Likert-type scales (where the respondent rates his/her agreement with a series of statements on an agree-disagree continuum). Such direct estimation methods are very commonly used, although they are prone to biased responding. Furthermore, the analysis of the data arising from their use is not always appropriately analysed.

Respondent bias can be reduced with comparative methods which use expert judgement to scale the value of each item response in advance. The result of this is that when the questionnaire instrument is used, the extent to which each response option represents the 'right' or 'wrong' answer is to some extent hidden from the respondent, thus making biased responding more difficult. A range of established comparative methods can be used for this purpose, and provide response values on an interval scale.

Instrument scoring

When instruments consist of multiple items that are all designed to measure the same attribute, combining these should increase reliability (see below). A scaling model is a procedure that allows weights to be devised for instrument items according to the level of the attribute of interest associated with the given item (Nunnally and Bernstein, 1994). Thus if vocalisation is an item in a pain measurement instrument, a scaling model can determine its contribution to that measurement, relative to other items designed to measure the same attribute within that instrument.

The simplest scaling model which uses expert judgement and statistical modelling to define item weights is the equally weighted model which assumes an equal weight for each of the items so that the total score for the instrument represents the number of items observed when the assessment is made. Other models use comparative methods (such as Thurstone's method, Guttman scaling and paired-comparisons technique) to determine a weight for each item, and the total score for the instrument can then be calculated by adding together the weightings for the items observed when the assessment is made (e.g. Morton et al, 2005). More recent developments have made use of Item Response Theory (IRT) to map item responses to an attribute continuum (e.g. Burnell, 2004). All of these techniques are appropriate for scaling items designed to measure a single attribute.

Other multi-item instruments – sometimes described as composite instruments – contain items intended to measure a complex construct consisting of a number of distinct attributes. These instruments are usually constructed by making deliberate (and justified) choices about which variables to include and about how these should be combined to generate instrument scores (Fayers, 2004). For example, a poor score on any one causal variable may be sufficient to predict pain or HRQL, so it may be more appropriate to use the item maximum or minimum as the score for a particular subscale or scale. For example, where several items relate to lesions, for example, the score for that part of the instrument might be better represented by the poorest score obtained from among those items rather than the average of their scores. Responses to composite instruments may be combined to produce a single score, or a pattern or profile of scores may be generated. In either case, the scoring methodology should be clearly justified.

Reliability

Reliability is a measure of whether an instrument can measure accurately and repeatedly what it is intended to measure. If an instrument is to be used by an independent observer, then inter-rater reliability should be sought; alternatively, an instrument's test-retest reliability can be estimated by examining the stability of responses when scores are not expected to change between administrations (e.g. Holton et al, 1998). If an instrument is valid then it is likely also to be reliable, but it may be highly reliable yet lack validity because it is measuring something other than that which it was intended to measure (Fallowfield, 1990).

Responsiveness

While reliability is an important attribute of an instrument, it is possible for an instrument to be reliable yet be unresponsive or poorly responsive to clinical change. The ability of an instrument to capture changes that are important (statistically and practically) has been termed its responsiveness. This is an essential requirement of the evaluative instruments – those designed principally to measure change over time (Guyatt et al, 1987) – which are required for most clinical uses.

Utility

A useful clinical instrument must not only be valid, reliable and responsive but also 'practical and easy to administer, score and interpret' (Landgraf and Abetz, 1996). The possibility of self-administration and literacy level required of respondents are also utility considerations.

Conclusion

A number of tools intended to measure pain and HRQL in animals have been developed in recent years, and range in complexity from a single VAS or NRS on which to record a global score for 'pain', to composite instruments developed using the psychometric methods described above (or similar) such as the instruments developed in Glasgow for measurement of canine acute and chronic pain.

A range of instruments are referenced or detailed later in this chapter (for acute pain assessment) and in Chapter 9 (for chronic pain assessment). Because of the time-consuming nature of instrument development and testing, veterinary practitioners should consider the suitability of existing instruments before embarking on the development of others.

For veterinary practitioners who want to improve the management of pain in their animals and are faced with a range of instruments to choose from, the key questions to ask when selecting an instrument for a particular purpose are listed below.

- What does the instrument intend to measure, and what is its validity for this purpose (e.g. consider how measurement goals were defined, how items were generated and selected, and with what population(s) and in what contexts the instrument has been tested?
- Do the response options seem appropriate and is the method of generating instrument scores from item responses fully explained and justified?
- What evidence has testing of the instrument provided for its:

- criterion or construct validity?
- reliability?
- responsiveness (for evaluative instruments)?
- utility?

If no suitable instrument exists, and a decision is made to devise a new instrument for a particular purpose, care should be taken to use best practices in instrument development and testing (and these should be published) before the novel instrument is used for clinical or research purposes.

References

Abell N, Springer DW and Kamata A (2009) *Developing and Validating Rapid Assessment Instruments*. Oxford: Oxford University Press.

Brown D, Boston RC, Coyne JC and Farrar JT (2007) Development and psychometric testing of an instrument to measure chronic pain in dogs with arthritis *AJVR* 68(6): 631-637

Burnell MG (2004) *The Scaling of Acute Pain in Dogs*. Thesis submitted for the degree of PhD, University of Glasgow.

Cook KF, Monahan PO and McHorney CA (2003) Delicate balance between theory and practice: health status assessment and item response theory (Editorial). *Medical Care* 41(5): 571-574

Fallowfield L (1990) *The Quality of Life: the Missing Measurement in Health Care*. London: Souvenir Press

Fayers PM (2004) Quality-of-life measurement in clinical trials – the impact of causal variables. *Journal of Biopharmaceutical Statistics* 14(1): 155-176

Flecknell P (2008) Analgesia from a veterinary perspective. *British Journal of Anaesthesia* 101(1): 121-4

Guyatt G, Walter and Norman G (1987) Measuring change over time: assessing the usefulness of evaluative instruments. *Journal of Chronic Diseases* 40(2): 171-178

Guyatt GH, Feeny DH and Patrick DL (1993) Measuring health-related quality of life. *Annals of Internal Medicine* 118: 622-629

Heilm- Bjorkman AK, Kuusela E, Lipman A, Markkola A, Saarto E, Huttunen P, Leppaluoto J, Tulamo R-M and Raekallio Ml (2003) Evaluation of methods for assessment of pain associated with chronic osteoarthritis in dogs. *JAVMA* 222(11): 1552-1558

Holton L, Scott EM, Nolan, AM, Reid J, Welsh E and Flaherty D (1998) Comparison of three methods used for assessment of pain in dogs. *JAVMA* 212: 61-66

Jensen MP (2003) Questionnaire validation: a brief guide for readers of the research literature. *The Clinical Journal of Pain* 19: 345-352

Johnston CC (1998) Psychometric issues in the measurement of pain. In: *Measurement of Pain in Infants and Children, Progress in Pain Research and Management Volume 10*, eds. Finley GA and McGrath PJ, pp5-20. Seattle: IASP Press

Landgraf JM and Abetz LN (1996) Measuring health outcomes in pediatric populations: issues in psychometrics and application. In: *Quality of Life and Pharmacoeconomics in Clinical Trials (2nd Edition)*, ed. Spilker B, pp 793-802. Philadelphia: Lippincott-Raven Publishers.

Loeser JD and Treede RD (2008) The Kyoto protocol of IASP Basic Pain Terminology. *Pain* 137(3):473-7

Molony VJ (1997) Comments on Anand and Craig, Pain 67 (1996) 3-6. *Pain* 70: 293

Morton CM, Reid J, Scott EM, Holton LL and Nolan AM (2005). Application of a scaling model to establish and validate an interval level pain scale for assessment of acute pain in dogs. *American Journal of Veterinary Research* 66(12), 2154-2166.

Nunally JC and Bernstein IH (1994) *Psychometric Theory*, pp 3-30. New York: McGraw Hill

Streiner DL (1993) Research methods in psychiatry. A checklist for evaluating the usefulness of rating scales. Canadian Journal of Psychiatry 38: 140-148

Streiner DL and Norman GR (2008) *Health Measurement Scales: a practical guide to their development and use, 4th Edition*. Oxford: Oxford University Press

Thomson W (1889) *Popular Lectures and Addresses vol. 1*, 'Electrical Units of Measurement', delivered 3 May 1883

Wemelsfelder F (1997) The scientific validity of subjective concepts in models of animal welfare. *Applied Animal Behaviour Science* 53(1-2): 75-88

Wemelsfelder F, Hunter TE, Mendl MT and Lawrence AB (2001) Assessing the 'whole animal': a free choice profiling approach. *Animal Behaviour* 62: 209-220

Wiseman-Orr ML, Scott EM, Reid J and Nolan AM (2006) Validation of a structured questionnaire as an instrument to measure chronic pain in dogs on the basis of effects on health-related quality of life. *American Journal of Veterinary Research* 67(11): 1826-1836

Wiseman-Orr ML, Nolan AM, Reid J and Scott EM (2004) Development of a Questionnaire to measure the effects of chronic pain on health-related quality of life in dogs. *American Journal of Veterinary Research* 65(8): 1077-1084

Yazbek KVB and Fantoni DT (2005) Validity of a health-related quality of life scale for dogs with signs of pain secondary to cancer. JAVMA 226(8): 1354-1358