THE DEVELOPMENT OF AN INSTRUMENT TO MEASURE CHRONIC PAIN IN DOGS

Margaret Lesley Wiseman-Orr

A thesis submitted for the degree of
Doctor of Philosophy

University of Glasgow

Division of Small Animal Clinical Studies
Faculty of Veterinary Medicine

May 2005

ABSTRACT

The population of geriatric dogs is growing, and with it the incidence of chronic and painful diseases, such as arthritis. There currently exists no valid, reliable and responsive instrument with which to measure disease progression and treatment effectiveness for such cases. The purpose of this study was to develop an instrument for use in a clinical setting to assess chronic pain in the dog.

Human chronic pain is recognised to be a complex, multidimensional experience that has a substantial impact upon quality of life (QoL). It is commonly assessed using structured questionnaires that measure the impact of pain upon a range of domains of QoL that are affected by changes in health state, often termed health-related quality of life (HRQL). Such instruments are developed using psychometric methods. Although in human medicine the self-report is regarded as the gold standard in the measurement of pain and of HRQL, for patients that are not capable of self-report, instruments are developed that depend upon someone very familiar with the sufferer to provide a proxy report on behalf of the patient.

Anecdotal reports of the literature suggested that canine chronic pain had an impact that was similar to that of human chronic pain on QoL, and highlighted the dog owner or carer as a potentially valuable observer of relevant behavioural disturbances. Consequently, this research applied the psychometric approaches of proxy human chronic pain and HRQL instrument development to the development of an instrument to measure chronic pain in the dog, using the owner to provide a proxy report. The development of the instrument followed established steps designed to ensure an instrument's validity: the identification of all domains relevant to the measurement of interest; generation of a pool of potential instrument items; selection of instrument items from the item pool, and validation of that

selection; design and pre-testing of the prototype instrument; field-testing of the instrument to establish its psychometric properties.

Domain identification was carried out through interviews with 47 owners of dogs suffering chronic pain. Potential items (descriptive terms) were generated using questionnaires completed by 165 dog owners. The items then selected for inclusion in the instrument were those most commonly used descriptive terms that adequately sampled the relevant behavioural domains previously identified. These domains and the items selected were subsequently validated by 12 veterinary practitioners and by 10 owners of dogs suffering chronic pain. The validated list of items was incorporated into a structured questionnaire, and this instrument was pre-tested using 26 dog owners. The finished instrument was then field-tested using the owners of 155 dogs who completed a total of 390 questionnaires prior to and during treatment at the University of Glasgow Small Animal Hospital and at a local Veterinary Practice, a majority of which dogs were suffering from chronic degenerative joint disease (DJD). A further 42 questionnaires were completed by the owners of 26 healthy controls dogs.

Factor analysis of the instrument responses for dogs suffering DJD revealed an interpretable 12-factor model, in which factors were interpreted as domains of canine HRQL: 'vitality', 'physical limitation', 'lethargy', 'anxiety', 'aggression', 'emotional upset', 'appetite', 'consistency of behaviour', 'mental disturbance', 'attention-seeking', 'sadness' and 'acceptance'. This analysis provided evidence for the construct (factorial) validity of the instrument, since responses to instrument items revealed an underlying structure that reflected the construct upon which the instrument was developed.

Scores were calculated for each of the 12 domains of HRQL identified by the factor analysis, and these were able to discriminate between dogs with chronic pain and healthy dogs on >86% of occasions. This provided additional evidence for the construct validity of the instrument, since scores obtained

with the instrument were able to discriminate well between groups known to differ on the attribute being measured.

Profiles of HRQL scores obtained for dogs with chronic pain were compared with those obtained for healthy dogs in a control group, and differences in these profiles were observed. An examination of changes in HRQL domain scores over time for individual dogs revealed that these scores tended to reflect clinical change in those individuals.

Additional validation studies of the instrument are required, along with an analysis of reliability and responsiveness. However, the results reported here suggest that the instrument developed in this study could, with appropriate refinement, be used to provide an HRQL profile of an individual dog diagnosed with a chronic and painful condition, which profile could be used to improve clinical decision-making on a day-to-day basis. The data obtained with such an instrument could further be used to facilitate the development of evidence-based therapeutic options for painful chronic diseases, and to help to define humane endpoints in order to reduce suffering.

The process described here offers a novel approach to the development of chronic pain and HRQL instruments for a range of animal species, and may have relevance for human chronic pain and HRQL instrument development.

This work is dedicated to my father

Prof. J. Stewart Orr

10 August 1930 – 21 October 2001

Always my loving guide and inspiration

And to my energetic, beautiful,
devoted and beloved Mum,

Jean W. Orr (née Williamson)

26 October 1931 – 18 January 2006

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

This work would not have been embarked upon without the encouragement of Professors Andrea Nolan, Jacky Reid and Marian Scott, and their support and advice throughout has been invaluable.

I am grateful to those members of the clinical, nursing and reception staff of the University of Glasgow Small Animal Hospital (UGSAH) who variously, often significantly, and always willingly assisted with this research. A number of veterinary practitioners outwith the University also participated in this work, and I thank them for their valuable contributions, particularly Ann Hutchison who took part in the field-testing of the instrument.

Many dog owners attending UGSAH participated in the early stages of the study, as did clients of the PDSA PetAid hospital, Shamrock Street, Glasgow. Later stages of the study involved clients of UGSAH and of Dalblair Veterinary Surgery, Ayr. This research would not have been possible without the participation of those dog owners, which absolutely essential contribution I acknowledge with gratitude.

I should like to thank the staff of University of Glasgow Media Services and secretarial staff in the Faculty of Veterinary Medicine for their assistance with taping and transcription of interviews. I should also like to extend my thanks to all of my colleagues in the Vet School who helped in numerous smaller ways and whose company made the days more pleasant.

Financial support was provided for much of this study by the Ronald Miller Postgraduate Scholarship and by Pfizer Animal Health, for which support I am very grateful.

Finally, and most of all, I thank my husband, Brian, and my children, Murray and Rosie, for their love and support, and for helping me to keep things in perspective, though good times and bad.

DECLARATION

I declare that the work presented in this thesis is my own unless otherwise stated and acknowledged.

Margaret Lesley Wiseman-Orr

Articles accepted for publication taken from the work presented in this thesis are:

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.

Scott E.M., Fitzpatrick J.L., Nolan A.M., Reid J. and Wiseman M.L. (2003) Evaluation of welfare state based on interpretation of multiple indices *Animal Welfare* 12: 457-468.

Wiseman-Orr M.L., Nolan A.M., Reid J. and Scott E.M. (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.

ABBREVIATIONS

AIMS	Arthritis Impact Measurement Scales
AVTRW	Association of Veterinary Teachers and Research Workers
CHD	Canine Hip Dysplasia
CHQ	Child Health Questionnaire
DEGR®	Gustave Roussy Child Pain Scale
DGQ1	Descriptor Generating Questionnaire 1
DGQ2	Descriptor Generating Questionnaire 2
DJD	Chronic degenerative joint disease
EDIN	Échelle Douleur Inconfort Nouveau-Né
EFIC	European Federation of IASP Chapters
EORTC	European Organization for Research and Treatment of Cancer
FA	Factor Analysis
FDA	Food and Drug Administration
FLIC	Functional Living Index–Cancer
GUVQuest (–Dog)	Glasgow University Health-Related (Dog) Behaviour Questionnaire
HAD	Hospital Anxiety and Depression scale
HRQL	Health-Related Quality of Life
HUI2	Health Utilities Index 2
IASP	International Association for the Study of Pain
IMMPACT	Initiative on Methods, Measurement and Pain Assessment in Clinical Trials
MID	Minimum Important Difference
MPQ	McGill Pain Questionnaire
MRC	Medical Research Council
NCCPC	Non-Communicating Children's Pain Checklist
NRS	Numerical Rating Scales
PCA	Principal Components Analysis
PDSA	People's Dispensary for Sick Animals (Shamrock St, Glasgow)
PICIC	Pain Indicator for Communicatively Impaired Children
PPPM	Parents' Postoperative Pain Measure
PCQL–32	Pediatric Cancer Quality of Life Inventory–32
QoL	Quality of Life
RMH–POQLQ	Royal Marsden Hospital Paediatric Oncology Quality of Life Questionnaire
UGSAH	University of Glasgow Small Animal Hospital
UGVS	University of Glasgow Vet School
VAS	Visual Analogue Scale
VRS	Verbal Rating Scale
WHO	World Health Organisation

Chapter 1

GENERAL INTRODUCTION AND REVIEW OF THE LITERATURE

1.1 Pain

Until the mid-20th century, human pain was perceived to be an entirely sensory experience. In the 1960s this simple model was replaced by a new conceptual model (Melzack and Casey, 1968) which presented pain as a multi-dimensional experience, consisting of three principal dimensions: a sensory-discriminative dimension that provided information on where, when, what kind and how much tissue damage had been caused; a motivational-affective dimension that was the disturbance of feelings of well-being to a greater or lesser degree of unpleasantness; and an emotional-evaluative dimension that was the psychological impact of the painful experience resulting in, for example, increased anxiety or depression or aggression, and so on. While there has since been some disagreement over the dimensional structure of the pain experience (Cleeland, 1989; Clark *et al.*, 1995) there has been a widespread acceptance that it is a complex and multi-dimensional one. This concept was encompassed in a definition of human pain first published in 1979 (IASP Subcommittee on Taxonomy, 1979) that has become internationally accepted within medical and scientific communities, and is here reproduced in part:

'Pain

An unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage.

Note: Pain is always subjective. Each individual learns the application of the word through experiences related to injury in early life. Biologists recognize that those stimuli which cause

pain are liable to damage tissue. Accordingly, pain is that experience which we associate with actual or potential tissue damage. It is unquestionably a sensation in a part or parts of the body, but it is also always unpleasant and therefore also an emotional experience…'

Despite its widespread acceptance, the IASP's definition of pain received severe criticism for its apparent discrimination against those pain sufferers who were unable to make a verbal report of their pain, such as infants and the cognitively impaired (Anand and Craig, 1996a; Craig and Badali, 1999; Cunningham, 1999) and also animals (Narsinghani and Anand, 2000).

For such sufferers it was argued that non-verbal behaviour should be regarded as a form of self-report (Anand and Craig, 1996b; Craig and Badali, 1999), a view supported by evidence that much communication between humans is non-verbal, particularly the communication of emotions (Eibl-Eibesfeldt, 1972; Argyle, 1987) and including the communication of pain (Prkachin and Craig 1995; Deyo *et al.*, 2004). In consequence, in 2001, on the recommendation of the IASP's Task Force on Taxonomy, a further 'Note' was added to the IASP definition of pain:

'Note: The inability to communicate verbally does not negate the possibility that an individual is experiencing pain and is in need of appropriate pain-relieving treatment.'

The general endorsement of the IASP definition, which emphasises the importance of the emotional component of the pain experience and has as its focus the individual's perception of pain, has led to significant advances in understanding the mechanisms, measurement and modulation of pain as a complex, multi-dimensional and wholly subjective experience.

For example, the physiological basis of pain perception is now much better understood than it was before Melzack and Wall's (1965) 'gate control' theory revolutionised thinking about its nature. Prior to that, 'specificity theory' had provided a simple model in which pain receptors generated impulses that were transported to a pain centre in the brain: an unsatisfactory model to

explain the complexity of the pain experience. Current understanding is of a much more sophisticated system, one that allows the signals indicating threat of or damage to the integrity of the body to be modulated, in order to intensify or reduce the perception of pain (Meyer *et al.*, 1994; Woolf, 1994; Fields and Basbaum, 1999).

Pain can be classified as nociceptive (or physiologic or 'normal') or pathologic (or pathophysiologic) (Devor and Seltzer, 1999; Muir, 2002). Nociceptive pain results when pain receptors (nociceptors) are stimulated by a noxious stimulus that may be chemical, mechanical or thermal, and when signals from these nociceptors reach a conscious brain via the spinal cord. Within the dorsal horn of the spinal cord, however, this transmission of pain information can be inhibited or amplified by means of spinal neuronal circuits and descending tracts from higher brain centres. Pain that is amplified by such central processes, or by peripheral processes such as the chemical sensitisation of nociceptors during inflammation (resulting in hyperalgesia and allodynia) has been described as pathologic or pathophysiologic pain. Such a category also includes central or peripheral neuropathic pain: pain that is caused by damage to or dysfunction in the nervous system itself.

The IASP definition of pain acknowledges that pain usually has a physical cause, but specifies that activity in the nociceptive pathways is not pain: pain is always 'a psychological state' (IASP Subcommittee on Taxonomy, 1979). Indeed, recent work using magnetic resonance imaging of the brain has revealed that regions of the brain activated during the perception of physical pain are similarly activated by an emotional trauma such as social rejection (Eisenberger *et al.*, 2003).

1.2 Chronic pain

In addition to describing pain by its mechanism, pain can also be described according to its temporal qualities: principally, pain may be either acute or chronic.

Acute pain is associated with tissue damage or the threat of this, and serves the vital purpose of rapidly altering behaviour in order to minimise further damage and to optimise the conditions (i.e. immobility and rest, and protection of damaged tissue) in which healing can take place; it stops when healing is complete, and has obvious survival value (Sternbach, 1981). Chronic pain persists beyond the expected course of an acute disease process (Russo and Brose, 1998), appears to be non-functional (Sternbach, 1981) and becomes highly debilitating (Craig, 1999).

The Darwinian survival value of acute pain requires that, in most circumstances, pain perception tends to override other cognitive activities and is difficult to ignore (Crombez *et al.*, 1999; Newton-John, 2003). Consequently, persistent pain tends to have a significant impact upon the psychology of the sufferer. Human chronic pain is often associated with fear, anger, anxiety or depression, all of which may be caused by and may in turn exacerbate the patient's pain (Wade *et al.*, 1990; Craig, 1999; Crombez *et al.*, 1999; Geisser *et al.*, 2000). In anticipation of pain, chronic pain patients may avoid physical and social activities associated with pain, resulting in physical deconditioning and increasing preoccupation with pain (Asmundson *et al.*, 1999; Crombez *et al.*, 1999; Newton-John, 2003). Chronic pain therefore can have a widespread impact on a patient's social and psychological as well as physical well-being. Consideration of these important psychological and social consequences has led to a more interdisciplinary approach to the problem of human chronic pain (Sternbach, 1981; Carr, 1999; Norton *et al.*, 1999; Keefe *et al.*, 2001).

However, although acute and chronic pain are now quite separately conceptualised, there is presently (Harstall and Ospina, 2003) no standard internationally accepted definition of chronic pain that identifies unequivocally a set of criteria, including duration of pain, which would permit the ready classification of pain as chronic. Chronic pain has been defined as pain persisting for more than 1 month beyond the resolution of an acute

tissue injury, pain persisting or recurring for more than 3 months, or pain associated with tissue injury that is expected to continue or progress (Beers and Berkow, 1999-2004). Many have considered only pain lasting longer than 6 months to be chronic (Russo and Broze, 1998; Worz, 2003). However, in a recent review published by the IASP (Harstall and Ospina, 2003), of 13 studies on chronic pain, 4 studies applied 6 months as the minimum duration of chronic pain, and the remaining studies used 3 months' duration to define chronic pain.

The IASP Task Force on Taxonomy sought to present definitions of terms and a classification of pain syndromes that would be widely accepted and used in human medicine. Their publication, *Classification of Chronic Pain* (Merskey and Bogduk, 1994), revealed the difficulty of providing a simple definition that would differentiate all relevant conditions. Three months was proposed as the most convenient point of division between acute and chronic pain for non-malignant pain, but not for malignant pain. For research purposes, pain lasting at least 6 months was suggested. In some circumstances, chronic pain is characterised by the persistence of pain when the process of repair is apparently ended, but for pain associated with chronic conditions such as osteoarthritis, such healing will not have taken place. On the basis that chronic pain is pain that persists beyond the normal time of healing, in practice such pain might have been suffered for less than a month. The Task Force concluded that 'given that there are so many differences in what may be regarded as chronic pain, it seems best to allow for flexibility in the comparison of cases and to relate the issue to the diagnosis in particular situations' (Merskey and Bogduk, 1994).

It is an unfortunate consequence of the plasticity of the human pain processing system that changes in the processing pathways can result in pain continuing even when the source of the original pain is removed (Woolf, 1994). Such chronic or recurrent pain is considered to be a disease in its own right, rather than just a symptom of disease, as recognised by the European

Federation of IASP Chapters (EFIC): 'although acute pain may reasonably be considered a symptom of disease or injury, chronic and recurrent pain is a specific healthcare problem, a disease in its own right' (European Federation of IASP Chapters, 2004). EFIC defines chronic pain as 'pain that lasts beyond the usual course of the acute disease or expected time of healing. It may continue indefinitely. Pain that is not relieved despite appropriate treatment is referred to as intractable pain' (European Federation of IASP Chapters, 2004).

1.3 Health related quality of life (HRQL)

Quality of life (QoL) is a term used in a variety of disciplines, with varied conceptualisations and definitions as a result of the highly abstract nature of the concept and the influence upon it of values and philosophical approaches (Dijkers, 1999). However, there is general agreement that QoL is a multi-dimensional construct that includes at least three broad domains, physical, psychological and social functioning, which the World Health Organisation (WHO) has identified as being essential to health (WHO, 1948).

Health-related quality of life (HRQL) is distinct from global quality of life in that it is primarily concerned with QoL domains that change as a result of ill health and medical interventions. HRQL is determined by 'the manner in which changes in health, particularly disease severity, co-morbid conditions, and treatment-related symptoms affect the similarity or discrepancy between expectations and experiences' relating to dimensions of well-being (Padilla *et al.*, 1996), and has been defined as 'a combination of health states and affective responses to problems in health status' (Theunissen, 1998). HRQL and health status are often used interchangeably in the field of human medicine. The difference between the two terms has been described as follows: 'quality of life, rather than being a mere rating of health status, is a uniquely personal perception, denoting the way that individual patients perceive and react to their health status and to other, non-medical aspects of their lives' (Gill, 1995). However, confusion about the definition of QoL has

resulted in overlap between measures of QoL and health or functional status and the domains they assess (Eiser and Morse, 2001).

Like pain, HRQL is conceptualised as a subjective perception, for which, as for pain, the patient should be the primary source of information (Sprangers and Aaronson, 1992; Sneeuw *et al.*, 1999).

1.4 Measuring human pain and HRQL

Pain is complex, but it has been argued that it should not be simplified in order to measure it (Chapman *et al.*, 1985). Unlike tangible attributes such as height and weight, which are relatively easy to measure, other attributes may initially exist only as theories that describe underlying 'constructs' proposed by the measurer until they can be confirmed. Some constructs are represented by a single attribute; others may consist of more than one. The debate over whether or not it is possible to measure something abstract of this kind is one that has not impeded progress in other sciences: chemists and physicists have been accustomed to hypothesising the existence of 'hidden' particles and forces and to testing those hypotheses experimentally (Schilhab, 2002).

Pain is currently conceptualised as a complex and subjective experience, of which the intensity of pain is only one attribute. The affective dimensions of pain have an important role to play in the overall experience of pain, yet, because they are subject to great individual variation and difficult to measure, clinicians and scientists have tended to pay much less attention to the affective dimensions of pain than to the sensory dimensions. Consequently, simple, unidimensional pain assessment tools such as verbal rating scales (VRS), visual analogue scales (VAS) and numerical rating scales (NRS) continue to be used to measure both human (Caraceni *et al.*, 2002, Luscombe and Williams, 2003; Ng *et al.*, 2003; Broström *et al.*, 2004; Fujita *et al.*, 2004; Turan *et al.*, 2004) and animal (Anil *et al.*, 2002; Hansen, 2003) pain. While such unidimensional instruments may be useful for measuring a single dimension of the pain experience, usually intended to be intensity, they are

not capable of capturing its complexity (Williams *et al.*, 2000; Clark *et al*, 2002). Although the clinician may believe that a unidimensional pain scale is measuring the intensity of a patient's pain, it has been shown that the patient's score may reflect either the emotional or the sensory dimensions of their pain experience, or a mixture of these, depending on the relative importance of each dimension for the individual at that time, but are often more influenced by the emotional than the sensory dimensions of pain (Williams *et al.*, 2000; Clark *et al.*, 2002; Knotkova *et al.*, 2004).

The success of an early multi-dimensional tool, the McGill Pain Questionnaire (MPQ) (Melzack, 1975), led to increasing interest in a multi-dimensional approach to measuring pain. Although the MPQ has been criticized for the complexity of its language and for its bias towards sensory aspects over pain's emotional component (Chapman *et al.*, 1985; Clark *et al.*, 1995) and for its focus on negative descriptors (Clark *et al.*, 1995), it signalled the beginnings of a willingness among researchers to embrace the complexity of pain in its measurement.

Chronic pain interacts in a complex way with a patient's social, psychological and physical well-being (Wade *et al.*, 1990; Asmundson *et al.*, 1999; Craig, 1999; Crombez *et al.*, 1999; Turk, 1999; Stroud *et al.*, 2000) affecting a patient's QoL (Becker *et al.*, 1997; Briggs *et al.*, 1999; Wirnsberger *et al.*, 1999; Katz, 2002). Recognition of this has led to the development of instruments to measure chronic pain through such impacts (Serlin *et al.*, 1995; Thomas *et al.*, 1996; Skevington, 1998; Bech, 1999; Briggs *et al.*, 1999; Penny *et al.*, 1999), and many of the instruments now used are concerned primarily with the way in which the chronic condition disrupts activities of daily living and alters HRQL (Schipper *et al.*, 1984; Ware and Sherbourne, 1992; Skevington *et al.*, 1997; Vallerand, 1998).

HRQL measures may be more responsive to clinical changes in patients with a chronic condition than are pain measures themselves (Skevington, 1998;

Gatchel *et al.*, 1999; Tugwell *et al.*, 2000). Consequently, HRQL has become an increasingly important focus of measurement of chronic pain in humans (Schipper, 1990; Lee and Rowlingson, 1996; Naughton *et al.*, 1996; Schlenk *et al.*, 1998; Gatchel *et al.*, 1999), with instruments such as the SF-36 (Ware and Sherbourne, 1992) and the WHOQOL (Skevington *et al.*, 1997) being used to assess both the impact of chronic pain (Becker *et al.*, 1997; Skevington, 1998) and treatment effects (Briggs *et al.*, 1999; Tugwell *et al.*, 2000).

Increased survival is often achieved by the aggressive use of potentially aversive treatment protocols that may adversely affect QoL both during and after treatment. For those whose care is palliative, the effectiveness of aggressive medical and psychological interventions may better be gauged by sequential HRQL assessments (Varni *et al.*, 1999).

Although HRQL instruments have been criticised for their focus only on negative aspects of health (Clark *et al.*, 1995; Ware, 1995), in human medicine, the ability of HRQL measures to assess the negative impact of medical interventions along with the beneficial ones, to assess outcomes other than cure where such an outcome is not possible, and to assess the affective component of chronic pain which may be the most significant to the patient, has led to them increasingly being used as measures of medical outcomes, which is an area of growing interest in human medicine (Ahmedzai, 1995; Bronfort and Bouter, 1999; Burgos-Vargas, 1999; Calaminus and Kiebert, 1999; Camilleri-Brennan and Steele, 1999; Feeney *et al.*, 1999; De Haes *et al.*, 2000; Patrick and Chiang, 2000).

Organisations such as the UK Medical Research Council (MRC) and the European Organization for Research and Treatment of Cancer (EORTC) consider that QoL should be a potential endpoint in clinical trials and that if QoL is not evaluated then justification for this should be provided (Fayers *et al.*, 1997). In the Unites States, the Food and Drug Administration (FDA) has recently encouraged testing of medications on paediatric populations during

drug development, with consequent demand for valid measures of HRQL for these populations (Matza *et al.*, 2004).

In response to this range of needs, the development and evaluation of HRQL instruments has increased exponentially in recent years (Garratt *et al.*, 2002). A human HRQL instrument could consist of a single question – 'How is your quality of life?' – but more often takes the form of a questionnaire consisting of a number of questions (usually called items) addressing different domains (or dimensions) of HRQL, including objective (what the individual can do) and subjective (the importance to the individual) assessments (Eiser and Morse, 2001). Measures of HRQL generally encompass four broad domains: physical/occupational status, psychological state, social interaction and somatic sensation (Schipper, 1990), addressing a wide range of 'facets of life' (Dijkers, 1999) such as mobility, physical activity, eating, sleeping, anxiety, alertness, depression and social activity as well as sensory pain (Fallowfield, 1990).

HRQL measures can be used to measure differences in the quality of life of different patients at a point in time (discriminative instruments) or to measure changes in HRQL within a patient over time (evaluative instruments). Some are specific, focusing on problems associated with, for example, particular conditions, particular populations or particular functions, and some are generic, sensitive to the impact of a wide range of diseases (Ware, 1995). Specific instruments may be more responsive to clinical change than are generic instruments (Wiebe *et al.*, 2003), although generic instruments can be valuable indicators of a range of effects of disease and its treatment on a patient's HRQL (Guyatt *et al.*, 1993; Graue *et al.*, 2003) and may be the only option when a patient is suffering from more than one condition (Eiser and Morse, 2001). In general, generic instruments have tended to be more refined and better validated than disease-specific instruments, although some well developed disease-specific instruments exist (Dijkers, 1999). Some examples of well-known HRQL measures include the Functional Living Index–Cancer

(FLIC) (Schipper *et al.*, 1984) and the Arthritis Impact Measurement Scales (AIMS) (Meenan *et al.*, 1982; Meenan *et al.*, 1992), both of which are disease-specific instruments, and the SF-36 (Ware and Sherbourne, 1992), and WHOQOL (WHOQOL Group, 1995; Skevington *et al.*, 2001), which are generic.

1.4.1 Measurement of multiple-attribute constructs

1.4.1.1 Psychometric and clinimetric methodology

Recent human pain instrument development has been firmly based upon the concept of pain as an abstract, multiple-attribute construct (Chapman, 1976; Wright and Feinstein, 1992). The psychometric methods established by psychologists and psychiatrists to measure such constructs, using formally-assessed structured questionnaires, have been adopted increasingly in the measurement of human pain, for example in the development of the Glasgow Pain Questionnaire (Thomas *et al.*, 1996) and the Non-Communicating Children's Pain Checklist–Revised (Breau *et al.*, 2000; Breau *et al.*, 2002a; Breau *et al.*, 2002b), which are used to measure a variety of pain types. The same approaches have been used to develop instruments to measure HRQL, since this is similarly intangible and may similarly be regarded as an abstract, multi-attribute construct.

Psychometry developed from psychophysics, which investigates the measurement by subjective judgement of physical phenomena that can also be measured by physical scales. Psychometrics adapted the psychophysical methods in order to measure, by subjective judgement, attributes for which no physical measures exist (McDowell and Newell, 1996). The psychometric strategy is usually aimed at finding and combining multiple items that measure a single attribute (such as anxiety or depression) and thereby increase the reliability of the measure. Because the multiple items are all intended to measure the same single attribute, psychometricians have sought to demonstrate that the items included in an instrument are relatively homogeneous. However, for measuring a multiple-attribute construct, the

multiple items need not be homogeneous: the most important consideration for such an instrument is to choose the best items to be included and emphasised, and, for measuring change, the items included should be those that are sensitive to the expected change. This is described as the clinimetric approach (Wright and Feinstein, 1992).

Many instruments designed to measure multiple-attribute constructs such as HRQL are based upon the sampling of variables of two quite different types (Fayers and Hand, 2002): indicator variables are those that may be considered to be indicative of an underlying attribute or attributes, whereas causal variables may be causal for the attribute(s) of interest. This is particularly true for disease-specific instruments for measuring HRQL, which, by comparison with generic instruments, often contain a high proportion of causal items such as troublesome symptoms or side effects associated with the disease and its therapies. An additional complication in HRQL measurement is that variables may be both causal and indicator, such as depression, which can be a result of pain and also have an influence on pain. Because misleading covariances may exist among causal items, where instruments contain both causal and indicator variables the psychometric quest for homogeneity among an instrument's items is not an appropriate approach (although it may be relevant at the level of subscales). That apart, both psychometric and clinimetric approaches to instrument development have important roles, the choice of method depending upon the purpose of the instrument and the items to be included, and both kinds of instrument can be developed with similar techniques (Wright and Feinstein, 1992).

The processes necessary for the creation of both psychometric and clinimetric instruments are well established (Streiner and Norman, 1995). They employ both qualitative and quantitative methods (Landgraf, 1999) and may be described in three phases. Phase 1 involves the specifying of measurement goals, the identification of the patient population, and the development of a pool of potential items for inclusion in the instrument. In Phase 2, suitable

items are selected from the item pool and that selection is subjected to expert validation. The validated collection of items is then incorporated into an instrument, with suitable consideration given to layout, response options provided for items, instructions to respondent and other details of administration. The resulting prototype is then pre-tested to ensure that the target respondent can use the instrument correctly. Phase 3 involves field-testing the instrument, in order to evaluate its psychometric properties (Streiner, 1993; Streiner and Norman, 1995; Juniper *et al.*, 1996).

The important contribution of the psychometric approach to instrument development is widely recognised (Cook *et al.*, 2003), and although researchers have been criticised for indiscriminate use of the methodology and for paying inadequate attention to the importance of the patient's individual values and preferences (Gill, 1995), these psychometric methods of instrument development have led to the creation of a number of instruments for the valid measurement of the subjective and multi-attribute constructs of pain (Thomas *et al,* 1996; Wheeler *et al.*, 1999; Debillon *et al.*, 2001; Breau *et al.*, 2002a; Ramelet *et al.*, 2004) and of HRQL (Eiser and Morse, 2001; Varni *et al.*, 2001; Garratt *et al.,* 2002; Matza *et al.* 2004).

The psychometric approach requires that instruments demonstrate the psychometric properties of validity, reliability and, usually, sensitivity to change, before being adopted for clinical use, and offers a range of methods for such evaluation. Criticism has been levelled at instruments developed with insufficient attention paid to such psychometric properties and to clinical utility (Abu-Saad, 2001). In a review of measures of QoL for children (Eiser and Morse, 2001), of a total of 43 instruments reviewed, the authors considered that only 3 generic and 2 disease-specific measures 'fulfilled very basic psychometric criteria'. However, the increasing emphasis on the importance of the scientific development and evaluation of new instruments (Coste *et al.*, 1995, Landgraf and Abetz, 1996) has led to improved reporting of the development of human instruments and their psychometric properties.

1.4.1.2 Validity

Validity is the most fundamental attribute of an instrument. It provides evidence that the instrument is able to measure the construct(s) that it was designed to measure. Validation of any HRQL instrument is an ongoing process, as new information is revealed for new conditions, new populations and the use of the tool in new settings (Landgraf and Abetz, 1996). Instrument developers should seek evidence for validity of 3 kinds: criterion validity, content validity (face validity, which is related to content validity, may or may not be sought) and construct validity (Streiner, 1993; Coste *et al.*, 1995; Johnston, 1998; Jensen, 2003).

1.4.1.2.1 Criterion validity

Criterion validity is the agreement of a new instrument (or parts of the new instrument) with some existing 'gold standard'. When no suitable gold standard exists, researchers use validation strategies established by clinical and experimental psychologists to provide evidence for content and construct validity.

1.4.1.2.2 Content validity and face validity

The content validity of an instrument depends upon the extent to which the attribute(s) of interest are comprehensively sampled by the instrument's items, and the appropriateness of each of the items to the measurement of interest. It is important that each item in the scale relates to what it is intended to measure ('content relevance') and that each of those areas of interest is represented by one or more items ('content coverage') according to its importance (Streiner and Norman, 1995). Content validity 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' who can confirm the appropriateness or otherwise of each item, and can supply any items deemed to be missing (Streiner and Norman, 1995).

An instrument that has face validity is one in which the items appear 'on the face of it' to be measuring what the instrument is intended to measure. This kind of validity does not improve the psychometric properties of an instrument, but it generally increases the instrument's acceptability to the respondent. However, in circumstances where there is a risk of biased responding, face validity may not be desirable (Streiner, 1993) and therefore may not be sought.

1.4.1.2.3 Construct validity

In psychiatry, the trait that is being measured is generally not itself visible, but is inferred from a variety of observations. The trait exists only as a hypothetical construct, which must be tested to provide evidence for the construct validity of the instrument (Streiner, 1993). Factorial validity is one kind of construct validity, which requires the statistical analysis of correlations between responses given to the items of an instrument. Groupings of items revealed by such analysis (that are also related on clinical or other grounds) are termed 'factors' and provide evidence for a factor structure underlying the data generated by the instrument. If this underlying factor structure fits the construct upon which the instrument was developed, then some evidence has been provided for the validity of the instrument and also for that hypothetical construct (Feinstein, 1987; Johnston, 1998). One of the drawbacks of this kind of testing is that both instrument and underlying construct are being tested simultaneously (Johnston, 1998).

Evidence for the construct validity of an instrument is also provided when the scores obtained with that 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). Selecting extreme groups is the easiest way to begin to establish such construct validity, where groups that should have high levels of the attribute are compared, using the instrument, to groups that should have low levels (Johnston, 1998).

1.4.1.3 Reliability

Reliability is a measure of whether an instrument can measure accurately and repeatedly what it is intended to measure, so that 'measurements of individuals on different occasions, or by different observers, or by similar or parallel tests, produce the same or similar results' (Streiner and Norman, 1995). If an instrument is to be used by an independent observer, then inter-rater reliability – when two or more observers concurrently applying the instrument to the same subject should provide similar scores – is a good indicator of the reliability of an instrument. Alternatively, an instrument's reliability can be estimated by examining the stability of responses when the instrument is administered on two occasions between which the scores are not expected to change: this is called test-retest reliability. This kind of reliability testing has been recommended for clinimetric scales that include causal as well as indicator variables (Fayers and Hand, 2002).

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). However, lack of reliability may limit the validity of an instrument (Streiner and Norman, 1995).

1.4.1.4 Responsiveness

While reliability is an important attribute of an instrument, it is possible for an instrument to be reliable yet be unresponsive to clinical change. A useful clinical instrument must be sensitive enough to detect differences in health status that are not only statistically important but are also important to the clinician or to the patient. This ability of an instrument to capture these kinds of change has been termed its responsiveness, which is considered to be an essential requirement of evaluative instruments – those designed principally to measure clinical change over time (Guyatt *et al.*, 1987).

There are a variety of statistical methods by which responsiveness may be evaluated, but none has become standard (Liang *et al.*, 2002). Variable results

have been reported (Wright and Young, 1997; Terwee *et al.*, 2003) for the responsiveness of the same health status instrument in the same applications when this was measured using different indices of responsiveness. Although it is not a universally held view, Beaton and colleagues (2001) proposed that responsiveness is not the inherent property of an instrument, but rather it is a property of an instrument's use in a particular context. This means that responsiveness can only be attributed to a particular application of an instrument, and not to the instrument itself, so that a study of responsiveness validates the application, but not the instrument.

In some cases, the most responsive scale may not be the best scale to evaluate efficacy of therapy – for example, if the scale is measuring dimensions that are reliably affected by the treatment, but are not of importance to the patient (Wright and Young, 1997). Researchers are therefore interested not only in the responsiveness of an instrument but also in whether changes measured with the instrument correspond with clinical changes that are familiar and meaningful to the clinician and to the patient. To describe such changes, one term and definition published is that of Guyatt and colleagues (2002), who proposed that 'the minimum important difference (MID) is the smallest difference in score in the domain of interest that patients perceive as important, either beneficial or harmful, and which would lead the clinician to consider a change in the patient's management'.

1.4.1.5 Utility

The utility of an instrument is a measure of its usefulness. 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). Even if a measure is valid and reliable, it may not have utility if it requires lengthy training, is time-consuming to administer or if scoring is complex (Streiner, 1993). The possibility of self-administration and the literacy level required of respondents are also utility considerations (Dijkers, 1999).

The utility of an instrument must be assessed in a particular setting with a particular population (Johnston, 1998). The importance of issues that may impact upon how readily a new instrument is utilised, such as the readability of a questionnaire, or how long a questionnaire takes to complete, has been highlighted, as has the need for developers and potential users to collaborate in order to ensure that such issues are addressed at the appropriate stage of instrument development (Landgraf and Abetz, 1996).

1.4.1.6 Choosing scales of measurement

Each item in an instrument is accompanied by an answer option or answer options, and an important consideration in instrument development is the choice of options to be offered to respondents. Response options may be dichotomous, or may be more complex, requiring some kind of measurement scale. Such a scale may offer nominal, ordinal, interval or ratio scale properties, each type offering different amounts of information and so different levels of measurement. The least information is provided by a nominal scale, which simply tells into which category a response falls. More information is provided by an ordinal scale, which ranks response options, providing information about how these relate to one another. Still more information is provided by interval and ratio scales. On an interval scale (such as a Celsius thermometer) response options are made on a scale of equal units, and a ratio scale (such as a centimetre ruler) has, in addition, a meaningful zero (Moore, 1991). The properties of a scale may determine the kinds of arithmetical and statistical operations that are appropriate to an analysis of the data they generate (Nunally and Bernstein, 1994).

If item responses are likely to lie on a continuum rather than be categorical, it is important to provide the opportunity for respondents to answer in this way to ensure minimum loss of information (Streiner and Norman, 1995), since 'the finer the distinction that can be made between subjects' responses, the greater the precision of the measure' (Bowling, 1991). Different types of scale are commonly used for the direct estimation of continuous variables,

including NRS, VAS, adjectival scales (with or without a VAS) and Likert scales (where the respondent rates his agreement with a series of statements on an agree–disagree continuum). Where a direct estimation scale offers a number of response options, there is evidence that around seven options tends to result in good reliability in scales in which people are asked to discriminate unidimensional stimuli or single attributes (Cichetti *et al.*, 1985; Preston and Coleman, 2000). This may be accounted for by the results of a study carried out in 1956 (Miller, 1956), which suggested that the human mind has a span of apprehension capable of distinguishing about 7 items (plus or minus 2), which implies a limit of about 7 on the number of categories that people are able to use in making magnitude judgements.

1.4.2 Self-report of human pain and HRQL, and the use of proxies

There is a footnote to the IASP definition of pain that states: 'pain is always subjective' (IASP Subcommittee on Taxonomy, 1986). Because of this subjective nature, in spite of their openness to biased reporting, self-reports are currently regarded as the 'gold standard' in assessing a person's pain (Melzack and Katz, 1999; McGrath and Unruh, 1999).

However, there are human sufferers who lack the necessary language skills or cognitive abilities to make such a report or to make a longer-term appraisal of events (Theunissen *et al.*, 1998; Eiser and Morse, 2001), for example, infants and those who are cognitively impaired. These individuals must rely on an observer to report on their behalf (McGrath *et al.*, 1985; Reid *et al.*, 1995; Buchholz *et al.*, 1998; McGrath *et al.*, 1998; Van Dijk *et al.*, 2000; Kappesser and Williams, 2002; Prkachin *et al.*, 2002; Stallard *et al.*, 2002). Recent work on methods of assessing pain in very young children and the cognitively impaired has focused on using the observations of caregivers to provide a report on behalf of the sufferer, for the measurement of acute (Gauvain-Piquard *et al.*, 1987; Gauvain-Piquard *et al.*, 1999) and prolonged (Debillon *et al.*, 1994; Debillon *et al.*, 2001) pain in infants, post-operative pain in young children (Chambers *et al.*, 1996; Chambers *et al.*, 2003), and pain in non-verbal,

cognitively impaired individuals (McGrath *et al.*, 1998; Breau *et al.*, 2000, Breau *et al.*, 2002a; Breau *et al.*, 2002b).

A joint position statement of the American Pain Society and the American Academy of Pediatrics states that 'observation of behaviour should be used to complement self-report and can be an acceptable alternative when valid self-report is not available' (reported in McClain, 2002). Facial expression has been the most comprehensively studied behavioural measure of pain in infants, and Franck and colleagues (2000) proposed that this should be considered the gold standard of behavioural response measures for pain in infants. However, the same authors, in their review of (mostly acute) pain assessment techniques for infants and children concluded that when using behavioural observation instruments health care providers consistently underestimated children's pain compared with the children's self reports, and that parents' ratings, although closer than those of nursing staff, also tended to underestimate pain (Franck *et al.*, 2000). It has been shown that proxies can be fairly accurate in assessing the more observable aspects of the pain experience of another person, such as physical functioning and impact on activities of daily living, but, using current methods, they are less good at assessing the more subjective elements, such as pain, feelings and thoughts (McPherson and Addington-Hall, 2003).

Like pain, human HRQL is similarly subjective but must be similarly assessed by proxy, where necessary, for adults (Sneeuw *et al.*, 1999; Councill *et al.*, 2001; Sneeuw *et al.*, 2002) and young children (Watson *et al.*, 1999; Armstrong *et al.*, 1999; Seid *et al.*, 1999; Varni *et al.*, 2001; Raat *et al.*, 2002).

An early review of studies in which self-report was compared with proxy report of HRQL, where a variety of methods of assessment were used, found that health-care providers tended to underestimate patients' QoL (as did lay individuals, such as spouses, involved in the care of the patient) and pain, though the ratings of lay caregivers tended to be more accurate when they lived in close proximity to the patient (Sprangers and Aaronson, 1992).

Another review of adult patient–proxy studies using well-known multi-dimensional HRQL instruments designed for self-report, or proxy-adaptations of these (Sneeuw *at al.*, 2002), found that judgements made by significant others (such as spouses) generally showed moderate to high levels of agreement with judgements made by patients, with mixed results reported for agreement between patients and health-care providers. A recent review of studies of HRQL assessment of children found that parents were the most common proxy respondents, that parents' reports were more accurate than those of doctors or nurses, and that approximately equal numbers of studies reported high or low parent–child agreement (Matza *et al.*, 2004).

A range of explanations has been proposed to account for discrepancies between parent and proxy report, including variations in 'internalised standards' for judging HRQL between adults and children (Theunissen *et al.*, 1998), children and parents having differing views on the impact of illness, and parental hopes, expectations, own stress or mental health (Eiser and Morse, 2001).

While self-report is often considered 'the gold standard' in the assessment of pain, it has limitations because the risk of biased responses is considerable. Hadjistavropoulos and Craig (2002) have argued that observation of automatic expressive behaviour may capture information about subjective states that is less open to purposeful distortion. Consequently, observational instruments can be used to measure another's pain not only when self-report is not possible but also when the credibility of self-report is in question. Hadjistavropoulos and Craig (2002) also argue that the degree to which a self-report pain assessment tool requires that attention be directed to the experience of pain, may itself affect that experience. A further problem associated with self-report is that language may be limited as a tool for the individual to convey the complexities of the pain experience (Craig and Badali, 1999).

1.5 Animal pain

Until recent decades, many were reluctant to accept that non-human animals might experience pain in a similar way to man. Despite the experts' acknowledgement of the importance of the 'feelings of animals' to an assessment of their welfare some twenty years earlier (Brambell, 1965), a definition of animal pain offered in 1986 excluded any reference to an emotional dimension that would make such pain something that was suffered: 'pain in animals is an aversive sensory experience caused by actual or potential injury that elicits protective motor and vegetative reactions, results in learned avoidance behaviour, and may modify species-specific behaviour, including social behaviour' (Zimmerman, 1986). Nevertheless, in the same paper the author goes on to state that 'it is essential to agree that animals can suffer' and suggests that 'pain is one cause of suffering'.

Growing societal concern for the welfare of animals soon led to the development of a clear concept of animal suffering as a highly unpleasant emotional response, usually associated with pain or distress (Kitchen *et al.*, 1987), and to a definition of animal suffering as 'an enduring negative emotional state associated with a perceived sustained threat to the integrity of the individual, helplessness and/or isolation from significant others' (Chapman, 1992).

Although scientists and philosophers continue to argue about the experience of pain in animals, and its similarity to the human experience of pain, there has been increasing acceptance that similarities in anatomy, physiology and pathophysiology between certain animals and man (Lamont *et al.*, 2000) support the hypothesis that those species may experience pain in a similar way. It has been suggested that of the three dimensions proposed by Melzack and Casey (1968), both the sensory-discriminative and the motivational-affective dimensions are likely to be part of the higher animals' experience of pain (Association of Veterinary Teachers and Research Workers, 1986). Dogs exhibit broadly the same physiological and behavioural responses to painful

stimuli as people do, and it is from these responses that we infer that a dog is suffering pain (Mathews, 2000; Muir and Gaynor, 2002). It is because of such similarities that dogs and other non-human animals have been used for research associated with human pain mechanisms and analgesia (Rollin, 1985).

Although the subjective nature of the pain experience renders impossible any scientific certainty about its perception (even about how it is perceived by another person), most scientists now consider that, unless proven otherwise, the morally correct stance is to assume that many non-human animal species may suffer pain in a similar way to man. Only a decade after Zimmerman's definition of animal pain was published, an alternative was proposed which recognised animal pain as a cause of suffering – as an aversive emotional experience as well as a sensory one – and one which had as its focus, as does the IASP definition, the perception of the individual:

'Animal pain is an aversive sensory and emotional experience representing an awareness by the animal of damage or threat to the integrity of its tissues (note, there may not be any damage); it changes the animal's physiology and behaviour to reduce or avoid damage, to reduce the likelihood of recurrence and to promote recovery. Non-functional pain occurs when the intensity or duration of the experience is not appropriate for the damage sustained (especially if none exists) and when physiological and behavioural responses are unsuccessful in alleviating it.' (Molony, 1997)

Not only do many now accept the likelihood that non-human animals may suffer pain as people do, some experts even argue that limited higher cognitive processes may mean that pain has the potential to cause greater suffering in non-human animals than it commonly does in people, since animals cannot 'understand' the cause of their pain (Rollin, 1985) or anticipate that it will be relieved (Robertson, 2002). This argument is not confined to non-human animals. An author writing on pain as it is experienced by a human infant argued that with minimal capacity to understand the meaning, significance and future of any painful event – the evaluative component of the

pain experience – the neonate and young infant's pain experience will be dominated by its sensory and affective parameters (Craig and Badali, 1999). Without the capacity to understand that suffering will not go on forever, it would seem possible that non-human animals with similarly functioning nervous systems and similarly limited cognitive capacities may experience pain in the same, potentially overwhelming, way that human infants do.

What is now widely accepted is that pain is not simply a sensory experience, and that it is the affective dimensions of the pain experience that make pain something that is suffered, rather than simply a useful sensory input (Craig, 1999). The fact that pain is suffered, however, may make the sensory input much more powerful and useful. The following description of pain affect highlights the value of suffering as an effective force for the promotion of survival and healing: 'pain also has a distinctly unpleasant, affective quality. It becomes overwhelming, demands immediate attention, and disrupts ongoing behaviour and thought. It motivates or drives the organism into activity aimed at stopping the pain as quickly as possible' (Melzack and Katz, 1992).

1.5.1 Animal chronic pain

Chronic pain is identified in Molony's definition of animal pain (Molony, 1997) as 'non-functional pain' that occurs 'when the intensity or duration of the experience is not appropriate for the damage sustained (especially if none exists) and when physiological and behavioural responses are unsuccessful in alleviating it'.

A widely used textbook of veterinary anaesthesia (Thurmon *et al.*, 1996) describes acute pain as 'the result of a traumatic, surgical, or infectious event that is abrupt in onset and relatively short in duration…generally alleviated by analgesic drugs'. By contrast, the authors' description of animal chronic pain (very similar to the accepted concept of human chronic pain) is lengthier and much less straightforward. It describes chronic pain as 'pain that persists beyond the usual course of an acute disease or beyond a reasonable time for

an injury to heal, or that is associated with a chronic pathologic process that persists or recurs for months or years'. By contrast with acute pain, which is described as a symptom of disease, Thurmon and colleagues consider chronic pain itself to be a disease; they recognise the important biological function of acute pain, whereas chronic pain is described as non-functional and detrimental to the patient; and they describe pain as 'a perception' and 'always subjective', for animals as it is for people.

Thus while there is consensus about the differences between acute and chronic pain, for both humans and non-human animals, one of which is the pain's temporal quality, there is no simple definition that would allow a condition to be classified as acute or chronic simply according to its duration. Individual studies must specify the criteria used when classifying pain as chronic. One recent study involving dogs suffering chronic pain (Muir *et al.*, 2004) defined chronic pain as pain of ≥1 month's duration, while another (Hielm-Björkman *et al.*, 2003) defined it as pain of >3 months' duration.

1.6 Animal quality of life

The term 'animal welfare' is a familiar one, but animal welfare is conceptualised in a variety of ways by scientists working in the field, as described by Duncan and Fraser (1997) and by Keeling and Jensen (2002). Largely, these fall into three categories, those who emphasise the importance of allowing the animal to lead a 'natural' life, those who emphasise the relevance to animal welfare of the biological functioning of the animal (animal growth, reproduction, longevity, etc.) and those for whom animal welfare is all about the subjective experience of animals (animal suffering, contentment, pleasure, and so on). Often, either of the latter two approaches would reveal the same assessment of animal welfare (but this is not always the case). Thus, because of the difficulty of measuring subjective states in others, existing measurements of animal welfare have tended to be made on those attributes that are more accessible, such as reduced life expectancy, impaired growth,

impaired reproduction, body damage, susceptibility to disease, adrenal activity and abnormal behaviour such as stereotypy (Broom, 1991). However, there may be occasions when such indicators of physical and mental condition are not directly associated with the subjective experiences of the individual. For that reason, attempts have been made to obtain, more directly, information about the subjective states of animals, by careful observation of behaviour and by experimental studies such as preference and motivation testing (Fraser and Matthews, 1997; Mench and Mason, 1997).

The term 'quality of life' has recently been used with reference to animal welfare and animal health (Taylor *et al.*, 1995; Clark *et al.*, 1997; Fraser *et al.* 1997; American College of Veterinary Anesthesiologists, 1998; Lund and Rocklinsberg, 2001; McMillan, 2003b; American Veterinary Medical Association, 2004; Watson 2004; Yearley *et al.*, 2004; Wojciechowska and Hewson, 2005), usually without definition. There is no widely accepted definition of the term animal QoL, but it has been equated with well-being as 'an individual's internal somatic and mental state that is affected by what it knows or perceives; its feelings and motivational state; the responses to internal or external stimuli or environments; individual variables, and phylogeny and ontogeny' (Clark *et al.*, 1997). A definition of animal QoL has been published by McMillan (McMillan, 2000). This also emphasises, as do current conceptualisations of human QoL, that animal QoL is multi-dimensional and subjective:

'Quality of life is a multi-dimensional, experiential continuum. It comprises an array of affective states, broadly classifiable as comfort-discomfort and pleasure states. In general, the greater the pleasant and lesser the unpleasant affects, the higher the QoL. Quality of life is a uniquely individual experience and should be measured from the perspective of the individual.' (McMillan, 2000)

Feeling or affective states (including pain) clearly have evolutionary value (Bateson, 1991; McMillan, 2001; McMillan, 2003a): it is proposed that they

have evolved to serve as mechanisms for encoding stimuli in such a way that pleasant or unpleasant feelings – of physical or emotional origin – will be associated with positive or negative influences (respectively) on survival and reproductive fitness. McMillan (2003b) suggests an array of factors that can contribute to animal QoL in this way, their importance varying from individual to individual: social relationships, mental stimulation, health status, food consumption, coping with stress, and control of environment or of relationship with environment. The same author recommends that assessment of an animal's QoL must be provided indirectly by a companion animal's closest human caregiver, and presents a very short questionnaire designed for this purpose (but for which no evidence of validity is provided) (McMillan, 2003b).

Fraser and colleagues (1997) have asserted that 'moral concern about the quality of life of animals arises because of the animals' capacity for subjective experience', but note that others have expressed the view that the subjective experience of animals, because it is not open to direct observation, 'falls outside the realm of scientific enquiry'. There has been increasing focus in animal welfare measurement on the individual animal's perception of its circumstances, and this mirrors the goals of those seeking to measure human pain and HRQL using the psychometric strategies already outlined.

1.7 Measuring an animal's subjective experience

1.7.1 Measuring animal pain

Animals are incapable of verbal self-report and so must, like non-verbal people, rely on an observer to assess their subjective experiences. There is evidence that simple rating scales such as VAS, NRS and simple descriptive scales (SDS) cannot be relied upon for the clinical assessment of acute pain in dogs (Hardie *et al.*, 1997; Holton *et al.*, 1998).

Behavioural disturbances have long been recognised as potential indicators of

the presence of pain in animals. Changes in demeanour, aggressiveness, submissiveness, fearfulness, restlessness, lethargy, activity, inquisitiveness, vocalisation, self-mutilation, appetite, drinking, urination, grooming and social behaviour have been asserted (Yoxall, 1978; Morton and Griffiths, 1985; Soma, 1985; Taylor, 1985; Association of Veterinary Teachers and Research Workers, 1986; Flecknell and Molony, 1997; Short, 1998; Mathews, 2000; Lester and Gaynor, 2000; Rutherford, 2002; Robertson, 2003; American College of Veterinary Anesthesiologists, 1995-2000), and the importance of making comparisons with the animals' pain-free behaviour is recognised (Morton and Griffiths, 1985; Taylor, 1985; Flecknell, 1985; Short, 1998). It has been suggested that, with chronic pain, by comparison with acute pain, changes in behaviour may be so gradual that they are apparent only to someone very familiar with the individual animal, such as the owner or carer (Flecknell, 1985; Brearley and Brearley, 2000).

The rating of pain-associated behavioural disturbances by a veterinary surgeon or a veterinary nurse has been the most active area of research in the assessment of acute pain in animals (Conzemius *et al.*, 1997; Holton *et al.*, 1997; Firth and Haldane, 1999; Cambridge *et al.*, 2000; Reese *et al.*, 2000; Holton *et al.*, 2001). Recent studies have highlighted the importance of the owner as contributor of information on behaviour changes to the assessment of chronic pain in dogs (Wiseman *et al.*, 2001; Gingerich and Strobel, 2003; Hielm-Björkman *et al.*, 2003). In two of these studies (Gingerich and Strobel, 2003; Hielm-Björkman *et al.*, 2003), questionnaires for owner response were tested as a suitable assessment method, and both studies found such an approach to be promising. In neither study were details provided of the methods used to generate the items that were included in the questionnaires tested. The authors recommended, for instrument refinement, the identification and inclusion of additional behaviours that may be sensitive to the impacts of chronic pain in individual dogs (Gingerich and Strobel, 2003) or in all dogs with chronic pain of particular cause (Hielm-Björkman *et al.*, 2003).

1.7.2 Ethology and anthropomorphism

An animal's behaviour offers the advantage of being available for study in a non-invasive and non-intrusive manner (Dawkins, 2004). In the earliest days of ethology – the study of animal behaviour as a science – every effort was made to describe and interpret the behaviour of animals in the simplest of terms, in accordance with Morgan's Cannon (1894): 'in no case may we interpret an action as the outcome of the exercise of higher psychical faculty if it can be interpreted as the outcome of the exercise of one that stands lower in the psychological scale'. This was a useful maxim in establishing ethology as a serious scientific discipline. Throughout this period, an anthropomorphic approach to understanding animal behaviour was to be strenuously avoided. However, in recent decades 'critical anthropomorphism' has been proposed as a useful tool for exploring potential similarities between man and other species and for predicting an evolutionary continuity of mental as well as physical characteristics between 'lower' animals and ourselves. Rasmussen and Rajecki (1995) stated that 'to the extent that people already use anthropomorphic models to understand one another, it seems no less legitimate to use anthropomorphism to predict or explain certain animal behaviour'.

It has even been argued by a number of authors that a rigid adherence to Morgan's Cannon may actually hinder progress in our understanding of animal behaviour: that interesting information may be lost by conceptually reducing an animal to a piece of clockwork machinery (Bateson, 1991), describing an animal's behaviour in purely mechanistic language (Christ, 1998) or emptying an animal's behaviour of its significance for the individual (Clark, 1990). Schilhab (2002), on this subject wrote: 'one is equally guilty of making a categorical mistake when wrongly denying counterparts of human mental states in animals when evidence clearly warrants drawing such conclusions'.

The traditional ethological methods of measuring behaviour are those in which an animal's apparently continuous stream of behaviour is regarded as a

series of discrete events: units of behaviour. Once identified, the frequency and duration of these units of behaviour is recorded and analysed to reveal patterns of behaviour that can provide information at a fairly high level about how an animal's behaviour is organised. This focus on the detail sometimes reveals features of the animal's behaviour that might otherwise be missed (Huntingford, 1984). An alternative to this approach is a rating method, in which an observer's rating of an individual's behavioural style (overall pattern of behaviour occurring in a variety of conditions and in complex social interactions) is formed. The human rater plays an active role in 'filtering, accumulating, weighting and integrating information over a considerable period of time' (Martin and Bateson, 1993).

1.7.3 Proxy assessment of a dog's mental state

It is easy to see how mental states, or feelings, linked to stimuli which have the potential to benefit or threaten animals, could have evolutionary survival value (Panksepp, 1998), the intensity of the feeling being related to the potential impact of the source of the stimulus (Johnston, 1999). Causes of unpleasant feelings may be physical (e.g. hunger, nausea, thirst, extremes of heat or cold, pain) or emotional (e.g. anxiety, fear, boredom). McMillan and Rollin (2001) note that 'feelings comprise all of life's pleasures, displeasures, enjoyments, miseries, and sufferings. For this reason, mental states are the only things in life that truly matter to animals'.

It is the unpleasant feelings associated with pain – pain's emotional component – that cause an animal to suffer. Since those feelings are the most important component of an animal's chronic pain experience, and it has been argued that pain should not be simplified in order to measure it (Chapman, 1976), it is those feelings that must be measured. But how can we measure something as abstract and subjective as an animal's feelings?

Dawkins (2004) has described an animal's behaviour, using Darwin's (1872) term, as 'the expression of the emotions'. Griffin (1992) asserted the

importance of recognising the value of our ability to make useful and correct inferences about the subjective feelings of other people by observing their behaviour, especially communicative behaviour, and that animal signals of this kind could be used to provide evidence about an animal's mental experiences. Recent work has demonstrated that naïve raters may be capable of identifying, with good agreement, subjective states in pigs (Wemelsfelder, 2000) and personality traits in dogs (Gosling *et al.*, 2003). It has been argued that it is legitimate to attempt to study such judgements scientifically (Dawkins, 1980; Bateson, 1991; Bekoff, 1994) and that a scientific exploration of the potential of the qualitative interpretation of behaviour for the assessment of an animal's mental state should be undertaken (Wemelsfelder and Farish, 2004).

Today's domestic dogs have evolved in association with man over thousands of years, suggesting that they have played an important role in human life since earliest times. In recent studies (Hare *et al.*, 2002; Miklósi *et al.*, 2004), researchers provided good evidence that in the process of their domestication dogs have been selected for a set of social–cognitive abilities that facilitate communication with humans, including the interpretation by dogs of human social cues and the generation by dogs of signals that humans are able to interpret. This high level of communication between man and dog makes the domestic dog a good candidate for a method of measuring pain that depends upon the communication of mental states.

In discussing subjective rating scales, Annett (2002) has argued for the value of using descriptive terms whose meaning is shared to describe observations and subjective experiences. The value of such communal lexicons is endorsed by others (Clark, 1998; Garrod, personal communication). The application to animals of terms usually used to describe human emotions, while anthropomorphic, does provide a vocabulary with shared meaning for the reporting of qualitative interpretations of animal behaviour.

1.8 The need for a valid instrument to measure chronic pain in the dog

The alleviation and control of pain are central to ensuring good animal welfare, and veterinary surgeons have a responsibility to provide these for the animals they treat (American Veterinary Medical Association, 2004; Royal College of Veterinary Surgeons, 2005). It has been proposed that pain be adopted as the fourth vital sign (temperature, pulse, respiration, pain) in veterinary medicine (Hellyer, 2002) but the fourth sign is very much more difficult to measure than the first three. A recent international workshop of experts in animal and human pain concluded that the creation of valid and acceptable instruments to measure animal pain is one of the most important tasks requiring immediate action (Paul-Murphy *et al.*, 2004).

In small animal veterinary practice there has been a marked change in the demographics of the pet population with an increase in the geriatric population of dogs and cats, resulting in more frequent presentations of painful, chronic conditions such as osteoarthritis and painful tumours (Lascelles and Main, 2002). The current trend in managing chronic osteoarthritis in dogs is towards using combinations of therapeutic agents accompanied by lifestyle and dietary management (Mama, 1999; McLaughlin, 2000) and oncology specialists must choose from a range of therapeutic options (Lester and Gaynor, 2000). In order to demonstrate the efficacy of selected treatments, the clinician must be able to assess, accurately, sensitively and reliably, clinical change in their patients.

1.9 The purpose and content of this research

The purpose of this research was to develop a tool to measure chronic pain, since this is presumed to have important welfare implications: pain is by definition an aversive experience – one that an individual would chose to avoid – and therefore one that will compromise welfare. Although there has been some recent research in this field (Hielm-Björkman *et al.*, 2003; Gingerich and Strobel, 2003), currently there exists no satisfactory, validated

tool for chronic pain assessment in dogs. The purpose of this research was to develop the first such instrument.

A background to the research has been provided in this chapter (Chapter 1), in which a brief but wide-ranging review of the literature described current concepts of human and of animal pain, and the complex and subjective nature of both of these experiences. It outlined the differences between acute and chronic pain, and introduced the concepts of human and animal QoL. The impact of chronic pain upon human QoL, and the relevance of QoL to the measurement of human chronic pain were described. The psychometric processes established for the development of human pain and HRQL instruments were detailed, and the shared requirement for proxy reporting of pain and HRQL for animals and people who cannot self-report was explained. Finally, current methods of assessing animal pain and of measuring animal behaviour were presented, along with recent arguments for the use of 'critical anthropomorphism' and the qualitative interpretation of animal behaviour in scientific studies.

The question about whether on not the experience of pain is similar in man and other animals is relevant to those who have an interest in animal welfare, whether this is in a professional capacity or not, and to those working in the medical field and using animals as human models. Pain has a sensory input and by definition (IASP, 1979; Molony, 1997) it is also unpleasant or aversive. The first two dimensions of pain proposed by Melzack and Casey (1968) for human pain – the sensory–discriminative and the motivational–affective dimensions – may therefore be considered to be the essential components of any experience that can be described as pain, in any species. Pain is not just what we feel but it is also, and probably more importantly, how it makes us feel.

Pain generates its emotional impact in a complex manner. First, there is the proposed motivational–affective dimension of pain, so-called because what is

felt has an affective dimension – pain's unpleasantness – that motivates the subject immediately to do something to avoid it. That immediate emotional dimension of pain cannot be separated from its sensory dimension, ensuring that the first indication of tissue damage is given a high priority for the immediate attention it requires (Wall, 1999).

In addition to the identified sensory–discriminative and motivational–affective dimensions of pain, a third dimension has been proposed for human pain – an emotional–evaluative dimension. This aspect of the pain experience was included in Melzack and Casey's multi-dimensional model, in which pain is influenced by 'cognitive or "higher central nervous system" activities (Melzack and Casey, 1968), such as remembered experience, anticipation and understanding. This cognitive influence can have the effect of reducing pain or of intensifying it, and may itself be influenced by that pain's impact upon the sufferer's psychological state.

Painful injuries and diseases can cause significant emotional distress in the form of fear, anxiety, anger and depression (Craig, 1999). There is evidence for a complex interrelationship between pain and its psychological impact, such that worsening pain increases such impact, and the psychological impact can intensify pain. The depression that is commonly suffered by those with chronic pain has been found to be a consequence of the extent to which increases in pain severity interfere with 'important life activities', so reducing positive social feedback and autonomy (Craig, 1999). It seems feasible that the lack of autonomy associated with uncontrollable pain could be a cause of depression in non-human pain sufferers. Some evidence for this has been provided in a method used to induce 'learned helplessness', a well-known animal model of depression, obtained by exposure of rats to unpredictable, inescapable, moderately painful electrical stimuli. Learned helplessness is associated with cognitive, motivational and emotional changes similar to those seen in depressed human subjects.

Many of the behavioural disturbances displayed by people suffering chronic pain (e.g. reduction in activity, appetite disturbance, irritability and social withdrawal) are interpreted as indicators of the debilitating psychological impact of their condition. It will be interesting, in this study, to compare the behavioural signs of chronic pain in dogs with those of chronic pain in people, in order to identify any similarities in these that might indicate that dogs are suffering the kinds of psychological impacts that are self-reported by people suffering chronic pain, and indicative of the substantial emotional impact of a chronic and painful condition. Such similarities would suggest that the experience of chronic pain for dogs may be very similar to our own experience of chronic pain, consisting not only of a sensory–discriminative dimension and a motivational–affective one, but including also some degree of emotional–cognitive involvement in the dog's experience of chronic pain.

As a result of the increasing recognition of the complexity of the relationship between chronic pain and its often significant psychological impact, for human chronic pain the best measure of clinical status and clinical change is now considered to be one that addresses that wider impact. One way of measuring that impact from the subject's perspective is simply to ask patients to take a global view by measuring their overall satisfaction with their circumstances – by asking 'what is your quality of life?'

There is much debate among philosophers, scientists and other interested parties concerned with human and animal welfare, about what exactly is 'quality of life', and what is required for an individual to have a good (or poor) QoL. It has been suggested that there are at least two distinctly different meanings of the term within two human fields in which it is widely used: in the social sciences QoL consists of 'objective living conditions and subjective satisfaction with them' while in medicine the same term describes 'the health related subjective well-being of the individual' (Birnbacher, 1999). For those who are interested in measuring animal welfare, the former conceptualisation may be useful in the measurement of farm animal welfare, where the principal

interest is in the effects of standard conditions on groups of animals. The latter conceptualisation may be the more appropriate for the measurement of the welfare of companion animals, where the focus is on the individual whose circumstances are likely to be unique.

However it is measured, in all fields in which QoL is a construct of concern, it is increasingly perceived to be a subjective evaluation. In recent years the World Health Organisation initiated a large-scale project with the purpose of developing an international quality of life assessment tool. The project began with clarification of the concept of QoL and involved agreeing a definition of QoL that was internationally appropriate and acceptable. The resulting definition of QoL was: 'an individual's perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns. It is a broad ranging concept affected in a complex way by the person's physical health, psychological state, personal beliefs, social relationships and their relationship to salient features of their environment' (WHO Group, 1995).

It seems reasonable and potentially valuable to adopt a similar conceptualisation for animal quality of life (QoL) as for human QoL, and a definition of QoL is proposed here that is intended to be relevant in any circumstances (including those of ill health):

Quality of life is the subjective and dynamic evaluation by the individual of its circumstances (both internal and external) and the extent to which these meet its expectations (innate or learned), which results in, or includes, an affective (emotional) response to those circumstances. (The evaluation may be a conscious or an unconscious process, with a complexity appropriate to the cognitive capacity of the individual).

Consequently, HRQL is conceived as the subjective evaluation of circumstances that include an altered health state and related interventions, so that while health may impact upon QoL in specific ways, it is considered that the simultaneous measurement of a subject's HRQL and QoL will provide

the same result, concurring with Birnbacher (1999) that 'there is no such thing as specifically medical QoL'.

Animals are thought to live, largely, in the present, with a limited capacity to consider circumstances that are not immediately perceived, either physically or temporally. People may include in their evaluation of their own QoL an awareness of circumstances that are physically or temporally remote, and the evaluation of QoL will undoubtedly include all circumstances of which the subject is aware. Nevertheless, the evaluation of human QoL is momentary, as it is for animals.

Since quality of life is an overall assessment made by the individual, it is possible to measure human QoL or HRQL by asking the simple question, 'what is your quality of life?' One of the difficulties of this approach is that individuals may be basing their assessments on very different things. For example, one patient may focus more on functional status, another on pain, and a third may try to weigh up each element and 'calculate' a synthesis of both. For the same patient, over time, pain may become more important than functional status, so that his overall assessment may be focused more on functional status on one occasion, and be focused more on pain on the next. Consequently, to allow for more meaningful comparisons between subjects or within subjects over time, instruments to measure human HRQL tend to be multidimensional, measuring separately the patient's evaluation of their physical, emotional and social well-being that are the requirements for good health (WHO, 1948). These three aspects may be considered as domains of QoL in which the effects of poor QoL may be apparent, or though which the impact of pain or disability may exert an influence on QoL. In other words, the variables measured in each of these domains may be indicator or they may be causal for QoL (Fayers and Hand, 2002).

Not only does addressing a number of QoL domains within an instrument ensure that all relevant variables are included in a measure, such an approach

also ensures that changes in one domain are not confused or obscured by changes in another when changes are not be uniform across all domains. This may be particularly important in the development of a QoL measure for dogs, because certain breed differences may be significant. For example, poor QoL may be more quickly reflected in reduced appetite for some breeds than for others that have a tendency to obesity and therefore may be expected to be highly motivated to eat (such as Labrador retrievers). Even within a breed, it is possible that individual differences, whether genetically or environmentally determined, will result in differences in the domains in which changes in QoL are reflected, or in the extent to which they are reflected in each domain

Because the behavioural disturbances anecdotally reported in the literature to be associated with canine chronic pain tended to be subtle, it had been suggested that such disturbances were therefore likely to be apparent only to someone very familiar with the animal. Consequently, it was considered that the dog owner or carer had the potential to make an important contribution to the measurement of canine chronic pain through its impact upon HRQL. This approach reflected that taken in human health measurement in which for HRQL and prolonged pain measures the preferred proxy respondent for non-verbal subjects is the person who knows the subject best, usually a parent or the spouse. The psychometric approach offers the greatest potential for the development of an instrument for use by a human observer that can be demonstrated to have the attributes that are crucially important for any instrument, of validity and reliability and, for certain uses, of responsiveness to change.

If QoL is conceptualised, as is pain, as an entirely subjective construct, the goal of its measurement must be to access that subjective perception. One of the potential criticisms that can be levelled at proxy instruments developed for human pain and QoL measurement is that such instruments are often developed from existing self-report instruments, which means that the proxy respondent is then required to make a fairly complex judgement that involves

'second-guessing' the responses that would be provided by the subject of the measurement if his or her self-report were available. An alternative approach taken in the development of other proxy instruments is to adopt an 'objective list' approach, in which the instrument contains a list of items that are based on external judgement of what is important for a good, or causal for a poor, QoL. This approach also has its difficulties, since what is important to one person at one time may not be to another person, or even to the same person at another time. Given the proposition that QoL is an individual appraisal of one's largely immediate circumstances, then it may not be desirable or even possible to provide an 'objective list' of items that would be appropriate in any but the most unsophisticated assessment. It would seem to be essential, for a true measurement of QoL, that the elements and any weightings included in that assessment are entirely left to the individual, and the instrument be designed to capture the output of that appraisal. Unfortunately, we cannot simply ask a dog to make a global assessment of its QoL. An owner, if asked the same global question, may will take an objective list approach in arriving at their proxy evaluation and consider that since their dog has a warm bed and two good walks a day, is well fed and receives immediate veterinary attention if required, it must have a good QoL, regardless of how the dog feels about those circumstances. This is a simplistic approach, since circumstances may be quite differently perceived by different canine subjects. For example, the opportunity of a long romp on a windswept beach is likely to be perceived in one way by an energetic young Labrador retriever and may be perceived entirely differently by an elderly Cavalier King Charles spaniel that has been raised as a lap dog.

An alternative to any attempt to list and rate the variables that may or may not be causal for QoL is simply to look for variables that appear to be indicators for QoL. One approach to this is to ask individuals to compare known groups (with different levels of the attribute in question) and simply ask them what they see. This approach is obvious to the veterinary instrument developer, who is not distracted, as is the human instrument developer, by the

content of self-reports provided by similar but not identical (e.g., older or less cognitively impaired) human populations.

When asking owners to report what behaviour they see, a decision must be taken about what form of report to seek. The pioneering work of Françoise Wemelsfelder has provided some justification for the use of a qualitative interpretation of animal behaviour as a means of obtaining information about the mental state of the animal (Wemelsfelder, 1997). Wemelsfelder and colleagues' sophisticated statistical analysis of naïve raters' interpretations of the subjective states of pigs (Wemelsfelder *et al.*, 2000, Wemelsfelder *et al.*, 2001) supported an hypothesis that owners would be able to provide reliable observations that would provide information about their dogs' mental states. The dog owners would undoubtedly, as did Wemelsfelder's observers, base their reports on an anthropomorphic interpretation of their observations, and use anthropomorphic terms to make those reports, since no other option would be readily available. However, in the case of Wemelsfelder's research, and in this study, such anthropomorphic interpretations and reports would subsequently be tested.

The term 'critical anthropomorphism was coined by Burghardt (1985) who argued that anthropomorphism was a legitimate approach to science if it was used to develop hypotheses that could be rigorously tested, and he proposed that critical anthropomorphism could use various sources of information including our perceptions, intuitions, feelings and identification with the animal in order to generate 'ideas that may prove useful in gaining understanding and the ability to predict outcomes of planned (experimental) and unplanned interventions' (Burghardt, 1991). The dog may be a particularly suitable animal with which to explore the potential of people to use their perceptions, intuitions, feelings and identification with the animal in order to identify a dog's behavioural expressions of affect. The dog has been bred over many years for the purpose of being a companion to, or of working with, man, and the importance of good inter-species communication in either

role is obvious. Other species may be expected not to reveal so much about their feelings to us as dogs might do. For example there may be advantages to prey species in being able to hide their vulnerability from predators. Communicating one's suffering is advantageous when it may elicit help from other social group members, but there may be another advantage that is not so immediately obvious. It has been argued (Bateson, 1991) that the debilitating effects of chronic pain could be advantageous in evolutionary terms, if they hasten the death of individuals for which there is little or no hope of recovery. It may benefit the genes of an individual in such circumstances if they make their condition obvious to predators, so that they may be picked off in preference to their fitter relatives who share their genes. It may therefore be valid to look for signs of chronic pain and poor QoL in a range of species other than dogs. However, even if behaviour indicative of suffering is obvious to the predators of a particular species, their behaviour may not be so open to naïve human interpretation if their behaviour is very different from our own.

The goal of this research was to develop an instrument to measure canine chronic pain and the proposed instrument was intended primarily to be a clinical tool. For clinical purposes, an evaluative instrument is normally required – one that can be used to measure change in an individual over time. Discriminative instruments, on the other hand, are used to measure differences between subjects at a point in time. Discrimination between subjects and evaluation of subjects need not require different instruments, but an instrument is unlikely to be equally effective in either role because it will generally have been designed with one or other purpose in mind. For that reason, it is important to identify at the outset the primary purpose of a novel instrument in order to achieve optimal efficiency for that identified purpose (Juniper *et al.*, 1996).

Specifying the measurement purpose will help the instrument developer to design appropriate development and testing protocols. The developer of a

discriminative instrument will focus, in item generation and selection, on identifying large and stable inter-patient differences, and will chose response options that achieve the discrimination goals: a crude measure will place subjects within a smaller number of categories than will a measure with greater discrimination. For example, a discriminative instrument may be required to discriminate only between subjects that differ greatly (e.g., those with mild disease and those with severe disease) or may be required to discriminate between subjects that differ only slightly (e.g., those with mild and those with very mild disease). The developer of an evaluative instrument will instead focus on items that are responsive to change and will select response options with sufficient gradations to register within-patient change. The developer of a discriminative instrument that will discriminate between those who have a particular disease and those who do not would omit any variables that were not specific to the disease since these may be common to both those who have and those who do not have the disease. The developer of an evaluative instrument to measure change in those who are suffering from that same disease might include those omitted variables, if they were variables that might be expected to reflect clinical change in the patients of interest.

Any instrument must be able to detect real differences (signal) above the random error associated with any measurement (noise). For an instrument to be able to discriminate between individuals, the variability in scores between subjects (the signal) must be sufficiently greater than the variability in scores within subjects (the noise). For an instrument to be used for evaluative purposes, the variability in scores within subjects who have improved or deteriorated (the signal) must be sufficiently greater than the variability in scores within subjects whose clinical condition has not changed (the noise). The former ratio is termed the reliability of an instrument and the latter its responsiveness. For an instrument to be suitable for evaluative purposes, it should offer both of these measurement properties.

Many instruments are used for either evaluation or discrimination, but the evaluative properties and the discriminative properties of an instrument must be considered separately when deciding whether or not it is appropriate for a particular use (e.g. Dempster and Donnelly, 2000).

The uses to which an evaluative instrument for the measurement of canine chronic pain through its impact upon HRQL could be put would include the monitoring of treatment effectiveness for individual dogs, whether such treatment was active or palliative, in order to guide treatment decisions, including decisions regarding the appropriateness of euthanasia. Such an instrument could also be used in clinical trials, to judge the effectiveness of one treatment compared with another, or with a placebo (if that was considered to be ethically acceptable). The current emphasis on evidence-based veterinary medicine, in which decisions on adopting, modifying or abandoning treatment methodologies are made according to peer-reviewed evidence, requires that robust measures of clinical impact be developed.

These are important welfare issues, particularly as treatment options increase, with some having associated unpleasant side effects, or significant negative impact in the short term for longer-term gain. Many clinicians and their clients would benefit from the availability of valid, reliable and responsive measures of pain and QoL as they attempt to identify the best course of action for the treatment of an animal that cannot speak for itself. With greater choice of treatment options, and increased affordability (at least for the growing numbers now choosing to take out veterinary insurance cover for their pets), has come increasingly demanding ethical decision-making in general as well as in specialist veterinary practice. An instrument that could be used with confidence to monitor clinical change in an individual, and to provide data that would facilitate the selection of treatments with known effectiveness and impact, should reliably inform such decision-making and so lessen the moral distress often suffered by those involved in such decisions, whether veterinary practitioner or client.

In this study, the hypothesis was made that a dog owner was able to report relevant behavioural disturbances, some of which were interpreted as the expression of the dog's mental state, and that such reporting and interpretation could be used to measure a dog's chronic pain through its impact upon HRQL. The accurate reporting of relevant owner observations and interpretations required that careful thought be given to how these would be made. For that reason, it was considered advisable to establish at the outset two important collections of information. First, a very clear understanding was required of the domains of behaviour in which owners observed behavioural disturbances associated with chronic pain in their dogs. Second, in order to describe such behaviour a comprehensive lexicon of terms would be required, with shared meaning within the dog owning community. In identifying these collections of information, it was intended to impose as little abstraction as possible upon the observations offered by dog owners. This initial work is described in Chapter 2.

Subsequently, these collections of information would be used to create a prototype instrument, following established psychometric methods for such development. A number of existing psychometric instruments appear to seek automatic reactions to items. This approach may be a means of accessing unconscious perception, since there is evidence that our affective reactions may be more influenced by stimuli that are unconsciously perceived than by those that are perceived consciously (Merikle *et al.*, 1998). The potential of this approach for obtaining valuable and unbiased information from owner observers was considered to be relevant to the instrument development that would be undertaken in this study. The process of instrument development is detailed in Chapter 3.

The prototype instrument developed would subsequently be field-tested with an appropriate population, to evaluate the construct validity of the instrument. The field-testing, and the results generated, are presented in Chapters 4 and 5.

Chapter 2

GENERATION OF POTENTIAL ITEMS FOR INCLUSION IN AN INSTRUMENT TO MEASURE CHRONIC PAIN IN THE DOG

2.1 Introduction

When developing an instrument to measure an attribute or construct of interest, the first step is to identify exactly what is likely to be relevant, and what is likely to be irrelevant, to such measurement (Juniper *et al.*, 1996). A review of the literature had identified that behavioural disturbances were likely to be relevant to the measurement of a dog's chronic pain (Chapter 1, pages 27-28). It had also identified the importance of the psychometric approach for the development of proxy instruments, based on behavioural observations, for the measurement of human pain and HRQL (Chapter 1, pages 11-21). It was hypothesised, therefore, that such an approach would be appropriate for the development of an instrument to measure chronic pain in the dog.

An established process for the development of structured questionnaires for the proxy measurement of human pain and HRQL begins with the identification of relevant behavioural observations and the generation of suitable potential items for the instrument. The content validity of an instrument is to a large extent dependent upon the source of the items from which it is constructed (Streiner and Norman, 1995; Juniper *et al.* 1996). Those with appropriate knowledge would include clinicians, patients, or, in the case of a proxy instrument, the potential proxy respondent.

For example, proxy human instruments such as the EDIN scale (Échelle Douleur Inconfort Nouveau-Né) (Debillon *et al.*, 2001) and the DEGR® Scale (the Gustave Roussy Child Pain Scale) (Gauvain-Piquard *et al.*, 1999) used the observations of nursing staff to identify behaviours relevant to the measurement of pain and to generate potential items; another, the

Postoperative Pain Measure for Parents (Chambers *et al.*, 1996; Chambers *et al.*, 2003) used the observations of parents for this purpose. In these cases, appropriate informants were those who were best able to report behavioural observations because their circumstances allowed them to make relevant observations over a suitable period of time.

Because the behavioural disturbances associated with chronic pain in animals are often subtle and gradual, it has been suggested that they may be apparent only to someone who is very familiar with the animal (Flecknell, 1985; Brearley and Brearley, 2000). Consequently, it was hypothesised that while the veterinary practitioner would be able to provide some relevant information about behavioural disturbances associated with chronic pain in the dog, it was likely that the reports of the dog owner would be more informative, both for the purposes of developing an instrument and for its subsequent application, because the dog owner is familiar with the dog's 'normal' behaviour and also because he or she is able to observe the dog in its usual environment over a suitable period of time.

In the field of human medicine, interviews with family members and professional caregivers were used to develop a valid checklist of behaviours that caregivers could use to identify pain in non-verbal, cognitively impaired people (Breau *et al.*, 2000), and to develop proxy paediatric QoL instruments (Varni *et al.*, 1998; Armstrong *et al.*, 1999). The semi-structured interview is a technique widely used in the social sciences, and established as a valuable method of obtaining data of a qualitative nature. The technique requires the interviewer to talk as little as possible while encouraging the interviewee to talk by asking prepared questions and using prompts in order to introduce a topic and then encourage its development. Interviews are usually taped, and various ethical considerations must be addressed (McCracken, 1988; Streiner and Norman, 1995; Gilchrist, 1999).

Successful communication requires an appropriate communal lexicon: words that have shared meaning for the community to which the participants belong (Clark, 1998). In questionnaire development, such a communal lexicon would be regarded as the most useful source of questionnaire items, and dog owners would represent such a community (Garrod, personal communication).

The format of questions and the design of a questionnaire are important considerations, since these can affect the validity of the instrument (Vaillancourt *et al.*, 1991). A well-designed questionnaire requires that questions be worded in a way that is clear and unambiguous and readily understandable, and the options provided for answering must be appropriate to the question and provide the opportunity for a respondent to answer accurately and sensitively (Payne, 1951; Streiner and Norman, 1995).

Consequently, the primary objectives of the work detailed in this chapter were the following:

- Interview veterinary practitioners and dog owners to identify behaviours observed to be disturbed by chronic pain;

- Generate a communal lexicon of terms used by dog owners to describe the behavioural disturbances that they associated with chronic pain in their dogs;

- Obtain additional evidence for owners' observations of behavioural disturbances associated with their dogs' chronic pain, and explore various methods of obtaining reports of such observations in a questionnaire format.

2.2 Identifying relevant behavioural disturbances through interviews with dog owners and veterinary practitioners

Preliminary interviews with dog owners and veterinary practitioners were used to obtain information about the kinds of behaviour changes observed with the onset of, or successful treatment of, a dog's chronic and painful condition.

Semi-structured interviews were used to establish with greater confidence the types of owner observations of dog behaviour likely to be relevant to the measurement of chronic pain, and the ways in which these were usually reported, by obtaining detailed accounts of observed behaviour changes from owners of dogs suffering chronic pain.

2.2.1 Preliminary interviews with dog owners and veterinary practitioners

2.2.1.1Materials and methods

Informal interviews were conducted with owners of selected dogs attending orthopaedic, oncology and soft tissue clinics in the University of Glasgow Small Animal Hospital (UGSAH). Selection of candidates for interview was made either prior to consultation (by the author, from information contained in clinical records) or during consultation (by the consulting clinician). Criteria for selection (by author or by clinician) were that candidates should be owners of dogs attending for a follow up appointment (not an initial appointment) and that those dogs should be suffering from arthritic conditions, anal furunculosis or painful tumours, or from chronic pain of another cause (as assessed by the examining clinician). Owners thus selected were asked in the course of the consultation if they had noticed any changes in behaviour since their dogs had become unwell or since treatment began. If so, they were asked if they would consent to be interviewed, following the consultation, about the behaviour changes they had observed. Those who agreed to participate were then given a short letter of introduction to the project, which explained the purpose and nature of the interview, and provided a name, address and telephone number for the project.

Owner interviews were held in private and lasted 10-20 minutes. Notes were handwritten by the author. In the course of the interviews, owners were asked to describe any changes they had observed in their dogs' behaviour since a painful condition developed or since treatment began. If observations were not readily forthcoming, attention was drawn to areas of behaviour in which

change might be expected (either because it was referred to in the literature or because it had been reported in interviews with other owners), such as sleeping, eating and drinking, exercising, playing, reacting to familiar or new people or situations, anxiety, scent marking, aggression, vocalisation, demeanour, consistency and self-cleaning behaviour. Following these interviews, details of diagnosis and treatment for each dog were obtained from hospital records.

Six veterinary practitioners were questioned about how they assessed chronic pain. Two of these veterinary practitioners were working in separate general practices, and four were specialists working in the UGSAH departments of orthopaedics, soft tissue, oncology and dermatology.

2.2.1.2 Results

Of the 26 owners selected as potential candidates for interview, 2 reported during their consultations that they had not observed any changes in behaviour since their dogs became unwell or since treatment began, and these owners were not invited to be interviewed.

Of the 24 dogs whose owners were interviewed, only 2 of the dogs were suffering from painful tumours and there were no cases of anal furunculosis. One dog had a chronic gastric problem. The remaining 21 dogs were attending the orthopaedic clinic.

Of the 21 orthopaedic cases, 3 of these were attending for fractures, and 1 for post-operative infection. The dog belonging to 1 owner was believed to have been in pain since coming into that ownership (rescue dog) so the owner was not able to report on how behaviour had changed compared with 'normal' behaviour for that dog. From hospital records, details of diagnoses and treatment for the remaining 16 orthopaedic cases were categorised by three orthopaedic specialists. Thirteen of the orthopaedic cases were thus categorised as chronic degenerative joint disease (DJD) (encompassing osteochondrosis, chronic osteoarthritis, cruciate failure, hip dysplasia, elbow

dysplasia, patella luxation and carpal hyperextension). The owners of this group of 13 mixed-breed dogs (8 male and 5 female) were interviewed either before treatment (2) or during/after treatment (11). Of the remaining 3 dogs, 1 was suffering from a spinal cord lesion, 1 from inflammatory polyarthropathy and 1 did not have a confirmed diagnosis.

During the period over which interviews were conducted, no cases with anal furunculosis were recruited, and only 2 cases with painful tumours. Consequently, only the group of dogs with DJD was sufficiently large for analysis to be undertaken without the risk of subject-specific physical limitations confounding the data.

The observed changes in behaviour reported in preliminary interviews by owners of dogs with DJD are detailed in Table 2.1. For comparison, behavioural disturbances reported by 9 other dog owners participating in preliminary interviews, but not the owner of the dog suffering post-operative infection, and not the rescue dog, are also shown in Table 2.1. The interviews with owners of dogs with DJD revealed owner-observed changes in behaviour across most areas of the behavioural repertoire. The range of behaviours reported disturbed by interviewed owners of dogs suffering from chronic and painful conditions other than DJD was very similar to that reported by owners of dogs with DJD, although most of the behaviours were reported to be disturbed with different frequencies in the two groups. The largest difference between groups was in the area of mobility, in which disturbances were reported by 84% of owners of dogs with DJD, and by only 44% of owners of dogs with other conditions.

In addition to reporting behavioural acts, 10 of the owners of dogs with DJD also described subtle aspects of their dogs' behaviour using terms and phrases that described styles of behaviour, using terms that described their dogs' attitude (e.g. 'half-heartedly', 'not so enthusiastic' and 'couldn't be bothered') or demeanour (e.g. 'miserable', 'a bit down', 'anxious' and 'distressed').

Table 2.1 Behaviours that owners (n=22) reported had altered in their dogs, 13 of which were suffering from chronic pain caused by DJD and 9 of which were suffering from chronic pain of other causes.

Behaviour	**Number of owners reporting behavioural disturbance in dogs with chronic pain caused by DJD (n=13)**	**Number of owners reporting behavioural disturbance in dogs with chronic pain of other causes (n=9)**
Mobility	11	4
Activity	8	5
Demeanour	8	4
Sociability	6	4
Playfulness	5	6
Vocalising	5	3
Anxiety	5	2
Daytime sleeping	5	4
Curiosity	4	4
Appetite	3	5
Aggression	3	1
Compulsive behaviour	3	2
Dependence	2	3
Fearfulness	2	2
Restlessness	2	2
Drinking	1	2
Scent marking	1	0

The signs of chronic pain described by the 6 veterinary practitioners are shown in Table 2.2. Four of the vets commented that behavioural disturbances were often more apparent to owners after their dogs had been successfully treated than they had been prior to treatment. The veterinary practitioners varied in the importance they gave to owners' assessments of their dogs' pain, although all considered that owners were able to provide some level of useful information, particularly when improvement was observed following treatment. Two vets commented on the necessity of questioning owners carefully about relevant behaviours in order to extract useful information. All of the veterinary practitioners interviewed felt that a tool for the more accurate assessment of chronic pain would be useful, largely for the clinical assessment of treatment effectiveness, but also for inter-institution communication and to assist owners and less experienced vets with decisions about treatment and euthanasia.

2.2.2 Semi-structured interviews with dog owners

2.2.2.1 Materials and methods

The owners of 25 dogs identified as suitable cases by veterinary specialists and general practitioners were interviewed either in UGSAH or in the People's Dispensary for Sick Animals (Shamrock St), Glasgow (PDSA). Criteria for case selection were that dogs should be suffering from a condition that the examining vet believed to be both chronic and painful, and that the owner should believe the dog to be suffering chronic pain. Sampling to redundancy was used to determine the number of interviews undertaken: interviewing ceased when it appeared that new interviews were not adding to the sum of information obtained from earlier interviews.

A standard ethics protocol ensured that interviewees understood that they were free to refuse to answer questions and to end the interview at any time, that the information they provided would be kept confidential and if published would be unattributed. An ethical statement to this effect (Appendix 1) was read to interviewees and a copy given to them.

Table 2.2 Signs of chronic pain reported by 2 veterinary general practitioners and 4 veterinary specialists in preliminary interviews.

Signs of chronic pain in the dog	Reported by
Lack of activity	both general practitioners; soft tissue, oncology and orthopaedic specialists
Change in demeanour	both general practitioners; dermatology, soft tissue, oncology and orthopaedic specialists
Dullness	both general practitioners
Depression	one general practitioner
Dullness of eye	both general practitioners
Reduced sociability	one general practitioner
Lack of appetite	one general practitioner; dermatology and oncology specialists
Abnormal movement or carriage	both general practitioners; dermatology, soft tissue and orthopaedic specialists
Abnormal responses	one general practitioner
Pain-related vocalising	both general practitioners
Panting	one general practitioner
Altered daytime/night time sleeping	one general practitioner; oncology specialist
Lack of interest	dermatology specialist
Reduced playfulness	one general practitioner; oncology and orthopaedic specialists
Dog 'not himself'	one general practitioner; dermatology and soft tissue specialists

Each interview began with the recording of biographical details (Appendix 2) followed by a 'grand tour question' – 'how can you tell that your dog is in pain?' – and then a series of 'planned prompts' asking about each type of behaviour in which change was expected: 'have you noticed any change in your dog's [specific behaviour]?' Additional information was elicited using 'floating prompts', for example, 'what do you mean by [owner's term]?' UGSAH interviews were conducted before PDSA interviews. All owners in UGSAH were asked the same questions; the script was slightly abbreviated for use in PDSA. Interview scripts are shown in Appendices 3 and 4.

The interviews, which lasted approximately 0.5h–1.5h, were taped using a Tandberg Educational TRC822 Audio Tutor, and later transcribed and analysed. Early audiotapes were transcribed by the University of Glasgow Media Services department onto microcassettes for secretarial transcription. Later transcription was done by the author direct from audiotape. All secretarial transcriptions were carefully checked by the author against the original audiotapes. Transcriptions were then analysed to provide quantitative and qualitative information about the kinds of behavioural disturbances observed by owners, and the ways in which these were reported.

2.2.2.2 Results

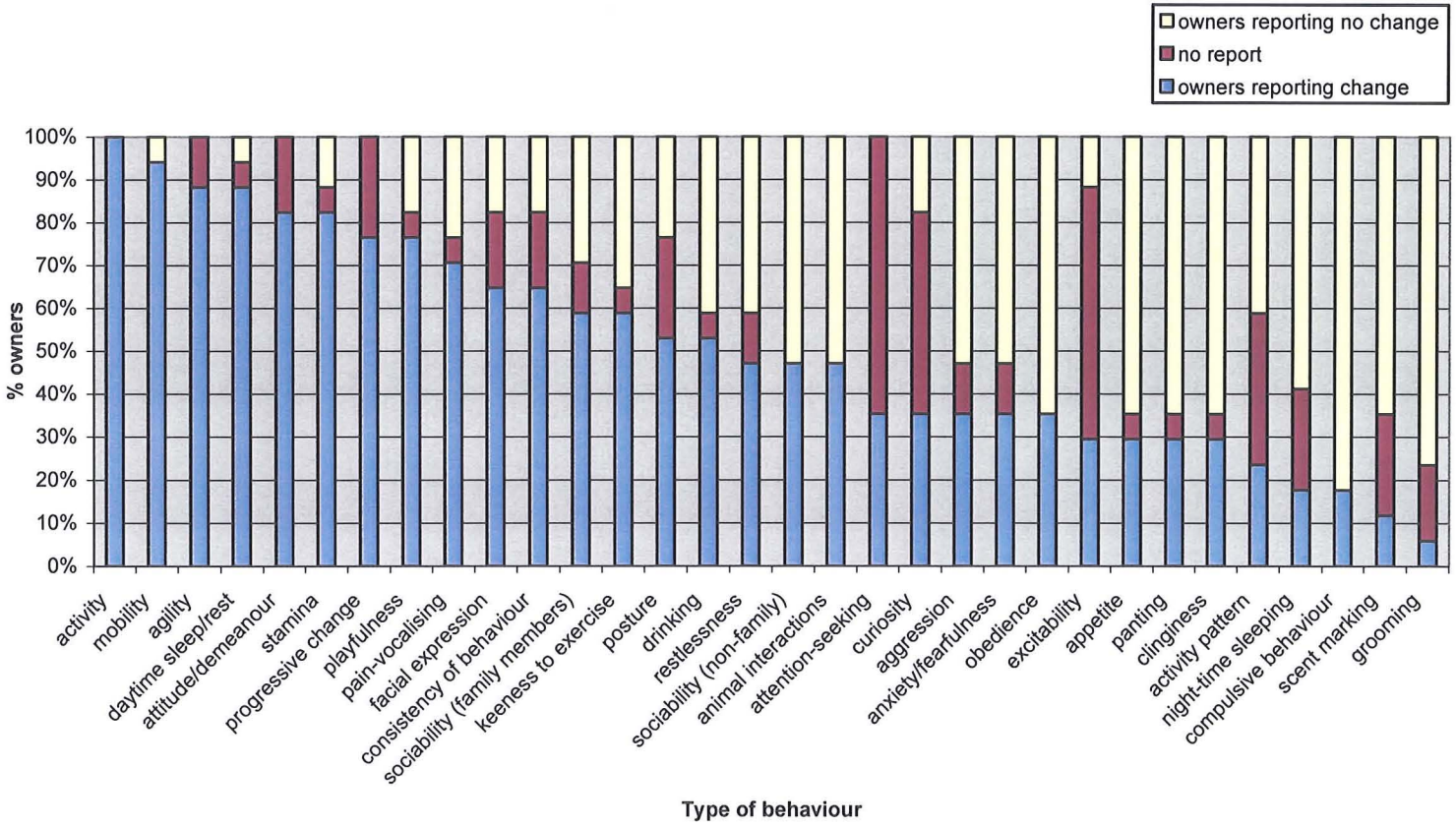
Of the 25 dogs whose owners were interviewed, subsequent confirmation of diagnoses revealed that 8 were suffering from a variety of conditions with potentially different condition-specific physical limitations. The remaining 17 dogs were suffering from conditions expertly categorised as DJD. The demographic details of these 17 dogs are shown in Table 2.3.

Disturbance of a total of 32 types of behaviour were reported by this group of owners, as shown in Figure 2.1. More than 75% reported changes in activity, mobility, agility, daytime sleeping/resting, attitude and demeanour, stamina, and playfulness, and reported behaviour changes as being progressive. Over 50% reported changes in pain-related vocalising, facial expression, sociability

Table 2.3 Demographic details of dogs suffering from DJD (n=17), the owners of which participated in semi-structured interviews.

Breed	**Sex**	**Age**	**Diagnosis**
Welsh Springer Spaniel	Male, neutered	16 months	DJD (hip replaced)
Crossbreed	Female, neutered	8 years	Cruciate failure
Labrador Retriever	Male	9 years	Osteoarthritis/cruciate failure
Golden Retriever	Female, neutered	10 years	Osteoarthritis/cruciate disease
Labrador Retriever	Male	10.5 years	Osteoarthritis
Labrador Retriever	Female	10.5 months	Osteoarthritis/ osteochondritis
German Shepherd	Male, neutered	5 years	Osteochondrosis
Labrador Retriever	Male	5 years	Cruciate failure/patella luxation
Crossbreed	Female, neutered	10 years	Osteoarthritis
Crossbreed	Male	14 years	Osteoarthritis
Golden Retriever	Male, neutered	13 years	Osteoarthritis
Labrador Retriever	Male	11.5 years	Osteoarthritis
Crossbreed	Female, neutered	13 years	Osteoarthritis
Crossbreed	Male, neutered	9 years	Osteoarthritis
Crossbreed	Female, neutered	14 years	Osteoarthritis
German Shepherd	Female, neutered	13.5 years	Osteoarthritis
Golden Retriever	Male	9 years	Osteoarthritis

Figure 2.1 Owner-reported (n owners=17) behaviour changes associated with chronic and painful orthopaedic conditions.



towards family members, keenness to exercise, drinking, and posture, and reported inconsistency in behaviour. Over 35% reported changes in restlessness, sociability towards people outside the family, interactions with other animals, curiosity, obedience, attention-seeking, aggressiveness, and anxiety or fearfulness. Other behaviour changes reported (by 18%-29%) included changes in appetite, excitability, panting, 'clinginess', pattern of daytime activity, compulsive behaviour, and night-time sleeping. Changed scent marking behaviour was reported by 2 owners, and changed grooming behaviour by 1 owner.

A qualitative interpretation of the transcribed interviews concluded that the 17 owners had confidence in their awareness of their dogs' behaviour, were capable of noticing gradations in behaviour, compared behaviour with their dogs' 'normal' (without pain) state, and interpreted some changes in behaviour as indicators of the mental state of their dogs (Table 2.4).

Owners frequently used descriptive terms or phrases that described a style of behaviour, for example, 'alert', 'clingy', 'depressed', 'dull', 'enthusiastic', 'irritable', often in conjunction with adverbs, prefixes or suffixes which expressed relation or degree, such as 'more', 'less', 'not as', 'extremely', 'very', 'a bit', 'not at all', un– (as in 'unhappy') and –er (as in 'slower') (Table 2.5). These descriptive terms or phrases described subtle aspects of behaviour – descriptions of attitude or demeanour – that owners were interpreting as expressions of their dogs' mental or emotional states.

2.3 Generating a collection of descriptive terms as potential instrument items

In order to generate a comprehensive collection of descriptive terms used readily by owners to describe the attitude and demeanour of their dogs in states of good health and in chronic pain, from which items could be selected for inclusion in an instrument, a series of descriptor generating exercises was carried out in UGSAH.

Table 2.4 Examples of extracts from semi-structured interviews with 17 owners of dogs with orthopaedic chronic pain, from which a qualitative interpretation of the data was made.

Examples of interview extracts from which a qualitative interpretation was made	**Qualitative interpretation**
'… this is a much more dramatic changeover than anything I've seen before with just the normal [age-related] change.' '… she would just suddenly perk up a bit.' '… it was such a marked change in him …' '… things that other people wouldn't notice.' 'Oh, I know her so well.'	Owners had confidence in their awareness of their dogs' behaviour:
'… slightly more excitable now …' '… a lot slower and stiffer …' '…she is a bit more careful now …' '… a bit more keen to get involved and interact ...' '… she gradually became slower…'	Owners were capable of remarking gradations in behaviour
'Where she used to be a companionable dog, I think she's more clingy at times and more distanced at times.' 'He was a dog who was always running about.' '… I wouldn't say he's as sociable now as he was before' '… he sleeps a lot anyway, but he was sleeping a lot more.' 'He's normally … quite a sociable kind of dog.'	Owners compared behaviour with their dogs' 'normal' (without pain) state
'… she is looking for sympathy rather than fun …' '… feeling sorry for himself …' '… he wants to be left alone.' '… she looks sad – you feel it, you feel it.' '… she was obviously very preoccupied with the pain.'	Owners interpreted some changes in behaviour as indicators of mental state

Table 2.5 Examples of descriptive terms and phrases used in semi-structured interviews by owners (n=17) of dogs with chronic orthopaedic pain.

Attribute described	Examples of terms and phrases used
Physical state or activity	(in-)active, aged, agitated, always on the go, boisterous, bouncy, can hardly move, (un-) comfortable, dancing, deliberate, didn't want to do anything, didn't want to play, energetic, exhausted, full of beans, full of life, gave up, hangy, healthier, hyperactive, jumped about, lame, leaping around everywhere, lethargic, lies about, limp, listless, lively (livelier), loved to run, mobile, not got the energy, playful, quiet(-er), relaxed, reluctant, restless, she will just… lie there, shuffling, sleeping all the time, slow(-er, -ed, -ing), sore, stiff(-er), tense, tight, tired, tried to play, was an effort, willing, with (had) difficulty
Appetite	eat like a horse, enjoy his food, good eater, greedy, not as good an eater, off her food, picky, wouldn't drink, wouldn't eat
Mental state or activity	affectionate, aggressive, alert, anxious, bored, bright(-er), can't be bothered, careful, cheerful, clingy (clingier), companionable, confident, confused, cry(-ing), daring, defensive, depressed, desperate, detached, distracted, distressed, docile, dull, eager, enthusiastic, excitable, excited, fearful, fed up, feeling sorry for himself, friendly, frightened, good-natured, groaning, growling, grumpy, half-hearted, hang-dog, happier, happy, inquisitive, interactive, interest(-ed), irritable, just wants to sleep, keen, laid-back, lazy, looking for attention, low, miserable, nervous, nosy, obedient, open, panicky, puts up with the pain, quiet, relaxed, remote, sad(-der), scared, screaming, sensitive, smart, snappy, sociable, stubborn, subdued, timid, tolerant, unsettled, wary, weeping, withdrawn, worried

2.3.1 Materials and methods

As a preliminary exercise the author carried out some informal questioning of approximately 50 owners of dogs visiting UGSAH over a period of 4 days. During this questioning owners were asked to describe their dogs' attitude and demeanour in simple terms (with a few, varying, examples of such descriptive terms given to aid understanding). Subsequently, descriptive terms were gathered in a more formal manner using questionnaires: DGQ1 and DGQ2.

DGQ1 Over a period of two months, this questionnaire was made available to all dog owners waiting to be seen in UGSAH. The questionnaire asked owners to suggest words that they would use to describe the attitude and demeanour of their dogs when well and when unwell. Alternatively, if attitude and demeanour did not change when unwell, they were asked to provide one list of words that would describe their dog in either state. Example descriptors were provided: affectionate, withdrawn, lively, and aggressive. No mention was made of pain, in order to reduce the risk of biased responses by owners who might have been reluctant to accept or admit the possibility of pain in their pet. DGQ1 is shown in Appendix 5.

DGQ2 This questionnaire was made available, over a period of approximately two months, to all dog owners waiting to be seen in UGSAH. Owners were asked briefly to imagine their dogs in chronic pain and to suggest words that they would use to describe their dogs in that state. They were then asked, on a second page, to select from an 'attitude and demeanour' list of terms provided, those terms that they would use to describe their dogs when in chronic pain. The 'attitude and demeanour' list was a list of terms owners had used to describe their dogs, which was derived from the data obtained from DGQ1 and from interviews. DGQ2 is shown in Appendix 6. Subsequently these owners were divided into two groups according to whether or not they had indicated that they were familiar with dogs in chronic pain. Group A were owners who reported that their dogs were at the time, or

had been in the past, suffering from a condition categorised by the author to be chronic and likely to be painful. Group B contained the remaining owners who may or may not have had direct experience of a dog suffering chronic pain.

The conditions from which dogs were suffering that were allocated to group A included osteochondritis dissecans (2 dogs), lameness of 6 months' duration (2 dogs), elbow damage of 5-6 months' duration (1 dog), carpal damage of 10 months' duration (1 dog), carpal problems of 5 weeks' duration (1 dog), osteoarthritis (3 dogs), ligament damage of 1 year's duration (1 dog), knee problems of 8 weeks' duration (1 dog), anal furunculosis (1 dog), hip problem of 2 months' duration (1 dog), possible lumbosacral disease of 2 months' duration (1 dog), and cruciate failure of 6 weeks' duration (1 dog).

2.3.2 Results

The preliminary informal questioning resulted in the collection of 466 descriptions in total, containing 184 different descriptive terms. The data indicated that, in general, owners had little difficulty in describing the attitude and demeanour of their dogs using simple, familiar terms.

DGQ1 Data were collected from a total of 93 owners. Sixty-three owners provided one list of words to describe their dogs when well and another to describe their dogs when unwell, or gave one of these lists without indicating clearly that attitude and demeanour were unchanged. Thirty owners each gave one list and clearly indicated that attitude and demeanour were unchanged whether their dogs were well or unwell.

After excluding 21 terms that were not descriptions of attitude or demeanour (such as 'expensive', 'fat' and 'barks') there remained 47 descriptive terms that owners used to describe the attitude and demeanour of a dog when either well or unwell, 64 terms that owners used to describe a dog when unwell, and 70 terms they used to describe a dog when well (Table 2.6).

Table 2.6 List of descriptive terms generated by questionnaire DGQ1, which asked owners (n=93) to provide a list of descriptive terms that they would use to describe the attitude and demeanour of a dog when well, terms they would use to describe a dog when unwell, and terms they would use to describe a dog in either state.

Well	active, affectionate, aggressive, alert, anxious, attention-seeking, attentive, aware, boisterous, bold, bouncy, bright, buoyant, busy, calm, cautious, cheeky, clumsy, comical, consistent, contented, curious, daft, demanding, eager, eating, energetic, excitable, exuberant, fit, friendly, fun-loving, funny, gentle, good-natured, greedy, happy, hungry, impatient, independent, inquisitive, interested, headstrong, joyful, keen, laid-back, lively, loving, loyal, mischievous, nervous, noisy, nosy, obedient, semi-obedient, outgoing, placid, playful, quiet, sensitive, smiling, sociable, soppy, stubborn, timid, vocal, waggy, unsociable, welcoming, well-behaved
Unwell	affectionate, agitated, aggressive, alert, anxious, attention-seeking, bored, clingy, crying, dependent, depressed, destructive, disinterested, distracted, dull, excitable, disobedient, friendly, frightened, gentle, greedy, guilty, hungry, inquisitive, irritable, lacklustre, lazy, lethargic, listless, lively, loving, methodical, moody, morose, nervous, panicky, placid, quiet, restless, sad, scratchy, shaky, sleepy, slow, slowed, stiff, sorrowful, strained, stressed, subdued, submissive, sulky, thirsty, tired, uncooperative, uneasy, unhappy, uninterested, unpredictable, unsettled, unsteady, weary, welcoming, withdrawn
Well/unwell (attitude and demeanour unchanged)	active, adaptable, affectionate, aggressive, alert, aloof, anxious, attention-seeking, boisterous, cheeky, child-friendly, confused, coughing, curious, devoted, energetic, excitable, friendly, fun-loving, good-natured, greedy, happy, happy-go-lucky, hungry, inquisitive, keen, lively, loving, minxy, nervous, nippy, noisy, nosy, obedient, outgoing, playful, quiet, rascally, relaxed, shy, single-minded, slowed, sooky, strong, stubborn, thirsty, timid

Excluding overlap, a total of 130 different descriptive terms for attitude and demeanour were collated. From this total, those selected for the 'attitude and demeanour' list to be included in DGQ2 (from which owners could choose terms to describe their dogs' attitude and demeanour) were relevant terms suggested by at least 2 owners (54 terms), together with 23 additional terms which although suggested by only 1 person in DGQ1 had also been suggested in the interviews with owners that have been reported in this chapter. This list of terms contained both positive (associated with well states) and negative (associated with unwell states) descriptors.

The following terms were suggested by at least two owners and were therefore included in the 'attitude and demeanour' list for DGQ2 (the terms 'hungry', 'lazy', 'panting', 'thirsty', 'stubborn', and 'welcoming', although suggested by at least two owners, were judged by the author to be describing something other than attitude and demeanour and were not included in the list): active, affectionate, aggressive, agitated, alert, anxious, attention-seeking, boisterous, bouncy, bright, calm, cheeky, clingy, comical, contented, curious, depressed, disinterested/uninterested, dull, energetic, excited/excitable, friendly, fun-loving, gentle, good-natured, greedy, happy, inquisitive, interested, irritable, keen, laid-back, lethargic, listless, lively, loving, mischievous, nervous, noisy, nosy, obedient, outgoing, placid, playful, quiet, restless, sleepy, slow(ed), sociable, subdued, timid, tired, unhappy, withdrawn.

In two cases, very similar words were considered for inclusion: 'disinterested' and 'uninterested' were considered to have the same colloquial meaning and only 'uninterested' was included in the list; 'excitable' was considered preferable to 'excited' and of these two suggested terms only the former was included in the list.

Twenty-three additional terms were included that had been suggested by fewer than 2 owners. These were included because they were judged by the author to be potentially useful descriptors because although suggested by only one person in DGQ1 all had also been suggested in informal questioning, in

preliminary or in semi-structured interviews. The terms were: bold, cautious, consistent, dependent, disobedient, eager, independent, lacklustre, moody, morose, panicky, relaxed, sad, sorrowful, strained, sulky, uncooperative, uneasy, unpredictable, unsettled, unsociable, weary, well-behaved.

The final list contained 77 descriptive terms, of which 39 were considered to be positive descriptors (descriptors more usually associated with well states) and 38 were considered to be negative descriptors (descriptors associated with unwell states), as shown in Table 2.7.

DGQ2 This questionnaire was completed by a total of 72 owners, of which 14 were allocated to group A and 53 to group B. One questionnaire from group A and two from group B were not correctly completed and were discarded. Groups A and B selected a total of 53 and 71 descriptive terms respectively from the 'attitude and demeanour' list, and collectively suggested an additional 131 terms and phrases to describe their dogs when in chronic pain. All but one ('mischievous') of the terms suggested by the owners in group A were also suggested by those in group B. These suggestions and selections formed an item pool from which could be selected the most appropriate items for inclusion in an instrument to measure a dog's chronic pain. This item pool is shown in Appendix 7.

2.4 Gathering additional information and testing question formats by means of an exploratory questionnaire

While the semi-structured interviews were being conducted, and while the descriptor-generating exercise using DGQ2 was ongoing, a questionnaire was devised by the author based on detailed descriptions of the kinds of behavioural disturbances reported in informal interviews. The purpose of this questionnaire was to obtain additional information from owners about the behaviour changes they observed in dogs suffering chronic pain and receiving treatment for a chronic and painful condition, while at the same time testing the suitability of a range of question types for the formal collection of information of this kind.

Table 2.7 Positive and negative descriptors selected for the 'attitude and demeanour' list incorporated in DGQ2.

Positive descriptors (descriptors associated with well states)	active, affectionate, alert, boisterous, bold, bouncy, bright, calm, cheeky, comical, consistent, contented, curious, eager, energetic, excitable, friendly, fun-loving, gentle, good-natured, greedy, happy, independent, inquisitive, interested, keen, laid-back, lively, loving, mischievous, noisy, nosy, obedient, outgoing, placid, playful, relaxed, sociable, well-behaved
Negative descriptors (descriptors associated with unwell states)	aggressive, agitated, anxious, attention-seeking, cautious, clingy, dependent, depressed, disobedient, dull, irritable, lacklustre, lethargic, listless, moody, morose, nervous, panicky, quiet, restless, sad, sleepy, slowed, sorrowful, strained, subdued, sulky, timid, tired, uncooperative, uneasy, unhappy, uninterested, unpredictable, unsettled, unsociable, weary, withdrawn

2.4.1 Materials and methods

Each page of a prototype version of the questionnaire contained a choice of descriptions intended to represent levels of expression of one type of behaviour over the previous week, along with a transition question about change in that behaviour since the dog's previous hospital visit (or since the dog was well) and a question about whether the behaviour tended to change from day to day. The opportunity was given on each page for respondents to write their own description of that kind of behaviour if none of the offered descriptions was suitable, and to add any other information they felt was relevant. Additional information about the meanings owners attached to various types of vocalisation was also sought. Also included was an opportunity for owners to propose terms they would use to describe their dogs' attitude and demeanour.

In order to establish that the questionnaire was acceptable to respondents, the prototype version (v.1) was completed (omitting the question about change since last hospital visit) by 16 dog owners recruited from among staff members of the University of Glasgow Vet School (UGVS). A revised version (v.2) was then pre-tested with 9 owners of dogs visiting UGSAH. Subsequent revisions resulted in a final version (v.3), completed by a group of owners whose dogs were attending UGSAH, for conservative treatment for arthritis, or for surgical treatment for a chronic and painful condition.

2.4.2 Results

The 16 owners (of 20 dogs) that completed v.1 were able to select a descriptive sentence for most types of behaviour for their dogs, but also offered more than 90 comments and suggestions for alternative descriptions, the content of which was reflected in revisions to the descriptions of behaviour included in v.2.

In the course of 1 week, 10 owners of dogs visiting UGSAH were approached to assist with the testing of v.2. Only 1 owner declined to participate. The 9 participating owners completed v.2 with the author present. The difficulties they had with completion resulted in a number of changes being made to the wording of questions, and to the format, in order to improve ease of use and enhance the quality of information obtained using the final version, v.3.

A sample page from v.3 is shown in Appendix 8. This questionnaire contained 28 A4 pages, each page concerned with one type of behaviour. A list of the behaviour types included is shown in Table 2.8. V.3 also contained 1 page for additional information about vocalisation, 1 page on which owners were asked either to suggest words to describe their dogs' general attitude over the previous week, or to select such words from a list of 77 terms describing 'attitude and demeanour' (the same list that was included in DGQ2), and 1 cover page bearing instructions, with space for entry of names of respondent and dog, and date of completion. V.3 was completed by 17 owners of dogs attending UGSAH for conservative treatment of arthritis. These owners of arthritic dogs completed from 1 to 4 questionnaires each in the course of treatment: 41 questionnaires were completed altogether.

Data obtained from this questionnaire confirmed that treatment of arthritis resulted in owners observing general changes, and some inconsistency, in a wide range of behaviours, as shown in Table 2.9. There were no types of behaviour, about which owners were questioned, in which no change was reported by any owner.

These data also revealed problems with various aspects of question design. All owners, on at least one occasion, anomalously reported no change in a type of behaviour from one visit to another, while selecting different descriptions of that behaviour on the two occasions, or even reported a change in behaviour that was the opposite of that expected by the owner's choice of descriptions on the two occasions.

Table 2.8 Behaviours included in v.3 of exploratory, information-gathering questionnaire.

Appetite	Enthusiasm for exercise
Playfulness	Exercise endurance
Drinking	Activity (mobility)
Fearfulness	Activity levels throughout the day
Social behaviour (towards family members)	Stiffness (mobility)
Social behaviour (towards people in general)	Stiffness (from lying)
Social behaviour (towards other animals)	Stiffness (agility)
Curiosity	Panting
Night-time sleeping	Anxiety
Working behaviour	Attention-seeking behaviour
Daytime sleeping	Compulsive behaviour
Aggression	Self-cleaning behaviour
Restlessness	Scent-marking behaviour
	Obedience
	Vocalisation

Table 2.9 Numbers of owners (n=17) of dogs receiving treatment for arthritis reporting some change from one questionnaire to the next ('general change') in the behaviours included in v.3 of exploratory, information-gathering questionnaire, and numbers of owners reporting these behaviours to change from day to day ('inconsistency').

Behaviour type	**Number of owners reporting general change**	**Number of owners reporting inconsistency**
Appetite	8	2
Playfulness	7	5
Drinking	4	1
Fearfulness	4	2
Social behaviour (towards family members)	7	2
Social behaviour (towards people in general)	6	1
Social behaviour (towards other animals)	6	0
Curiosity	5	2
Night-time sleeping	6	1
Working behaviour (None of the dogs were working dogs)	0	0
Daytime sleeping	9	4
Aggression	6	1
Restlessness	12	5
Enthusiasm for exercise	7	4
Exercise endurance	11	4
Activity (mobility)	9	5
Activity levels throughout the day	8	4
Stiffness (mobility)	13	7
Stiffness (from lying)	9	4
Stiffness (agility)	10	4
Panting	4	2
Anxiety	3	2
Attention-seeking behaviour	5	2
Compulsive behaviour	7	5
Self-cleaning behaviour	3	1
Scent-marking behaviour	2	1
Obedience	6	1
Vocalisation	8	1

On 54 occasions an owner wrote his or her own description of behaviour, because none of those offered described his or her dog well. Of the 17 owners in the group, 13 took this option on at least one occasion.

2.5 Discussion and conclusions

The complexity with which chronic pain is currently defined has been reported (Chapter 1, pages 3-6 and 24-25). The lack of a simple and widely accepted definition would have made it difficult to identify unequivocally, from among all dogs presenting in the clinical setting, those dogs that could be described as suffering 'chronic pain'. Consequently, it was decided at the beginning of this study to focus on dogs with a clinical diagnosis of one of a number of specific conditions widely believed within the veterinary community to be both chronic and painful. These conditions were DJD, anal furunculosis, chronic otitis externa and painful tumours.

A review of the literature had identified the dog owner as a potentially valuable informant for the purposes of establishing the domains of behaviour in which disturbances are observed in dogs with chronic pain. Because they are currently involved in making assessments of chronic pain in their patients, veterinary general practitioners and specialists were also regarded as potentially valuable informants for this purpose. The results of preliminary and semi-structured interviews with dog owners, and interviews with veterinary general practitioners and specialists, reported in this chapter, provided evidence that chronic pain caused by DJD in the dog has an impact upon a wide range of normal behaviours, and that these impacts may be subtle and of gradual development. These findings agreed with the anecdotal reports of the literature regarding signs of chronic pain in dogs (Flecknell, 1985; Soma 1985; Taylor, 1985; Mathews, 2000; Lester and Gaynor, 2000; Rutherford, 2002; Robertson, 2003). A similar impact on behaviour, described as a 'shutdown in activity' has been described in children whose pain is long-lasting (Gauvain-Piquard *et al.*, 1999) and in adults with chronic pain (Wall, 1979).

Therefore, in contrast to the way in which acute pain affects an animal's behaviour, which is in an abrupt manner and involves distinctive new behaviours such as flinching, guarding or vocalising (Holton *et al.*, 2001), which may be very obvious to an observer, chronic pain usually has the effect of gradually reducing the expression of a dog's 'normal' behaviour – the behaviour expressed when the dog is not suffering chronic pain – which may be less obvious to the observer. A similar problem has been recognised by those seeking to measure prolonged pain in human neonates, whose blank facial expression and reduced body movements often fail to be recognised as indicators of pain (Debillon *et al.*, 2001).

The findings that dogs with chronic pain show similarly diverse behaviour changes to those that are reported in human beings, such as changes in activity, sociability, aggression and appetite, supported a similarity between dogs and humans in the widespread impact of chronic pain on many aspects of QoL. It was therefore considered relevant to compare the domains of behaviour in which dog owners reported disturbances, with the content of existing instruments that have been validated for the measurement of human chronic pain and HRQL.

It was apparent, for example, that the range of behavioural disturbances reported by dog owners interviewed in this study reflected a number of the domains addressed in the well validated generic human HRQL instrument, the SF-36 (Ware and Sherbourne, 1992): physical functioning, role limitations because of physical health problems, bodily pain, social functioning, general mental health (psychological distress and psychological well-being), role limitations because of emotional problems, vitality (energy and fatigue), and general health perceptions. It also reflected some of the QoL domains addressed by the disease-specific Arthritis Impact Measurement Scales (Meenan *et al.*, 1992), validated for the measurement of the impact of chronic and painful orthopaedic disease in human patients on mobility, physical activity, dexterity, household activity, activities of daily living, anxiety,

depression, social activity and pain, satisfaction with health, function, and the patient's priorities for improvement.

Proxy instruments that have been developed to measure human pain and HRQL, and have been validated to some extent, address a similarly wide range of physical, psychological and social domains. For example, the Royal Marsden Hospital Paediatric Oncology Quality of Life Questionnaire (RMH-POQLQ) (Watson *et al.*, 1999) includes 70 items addressing functional status, physical symptoms, progress in school, emotional status, social functioning, cognitive functioning and behavioural problems. The Miami Pediatric Quality of Life Questionnaire (Armstrong *et al.*, 1999) contains 56 items within the domains of self-competence, emotional stability and social competence. The parent version of the PedsQL™ (Varni *et al.*, 1999) contains 8 items that address physical health and 15 addressing domains of psychosocial health. The DEGR® Scale, developed to allow nurses to grade prolonged pain in young cancer sufferers aged 2-6 years (Gauvain-Piquard *et al.*, 1999) contains 15 items concerned with signs of depression and anxiety and with behaviours specific to pain such as unnatural postures, pain avoidance and guarding. The Parents' Postoperative Pain Measure (PPPM), designed for parents to assess pronged post-operative pain in young children (Chambers *et al.*, 2003) contains 15 items addressing a range of behaviours including vocalisation, willingness to play, anxiety, vitality, appetite, social withdrawal and attention-seeking behaviour. The Non-Communicating Children's Pain Checklist (NCCPC) designed for caregivers to detect pain in non-communicating children (Breau *et al.,* 2000; Breau *et al.,* 2002a; Breau *et al.,* 2002b) contains 31 items within the subscales of vocal behaviour, eating/sleeping, social/personality, facial expression, activity, body and limbs, and physical signs.

In addition to evidence for chronic pain's widespread impact on dog behaviour, the interviews with owners also provided some evidence that dog owners were able to report relevant, subtle behavioural disturbances with

some confidence and often compared behaviour with the 'normal' behaviour of the individual dog. This provided support for the hypothesis that dog owners were potentially a valuable source of information about relevant behaviours, both for the development of an instrument to measure a dog's chronic pain and as a proxy respondent for this purpose, in accordance with the proposal that only someone very familiar with an individual animal, such as the owner or carer, would be able to provide relevant information for the assessment of animal chronic pain (Flecknell, 1985; Brearley and Brearley, 2000).

Work on assessing canine personality has recently made use of the technique used in some human personality assessment whereby judgements are made by informants who are well acquainted with the target individuals (Gosling *et al.*, 2003). In such assessment of behaviour an overall rating is given rather than a report of specific acts, durations and frequencies, which is the traditional ethological approach. The value of a 'rating' approach of this kind for revealing subtle information about an individual's style of behaviour has been recognised by ethologists (Feaver *et al.*, 1986; Martin and Bateson, 1993). Traditional ethological studies require that an observer has knowledge of and skills in recording methods, and is able to devote a considerable amount of time to observations. While it was considered inappropriate to ask the dog owner to undertake such a study, it was recognised that a simplified rating approach might facilitate owner reporting of relevant behavioural disturbances.

Interviews with veterinary specialists and general practitioners confirmed an influence of owner reports of behaviour upon the clinician's assessment of treatment effect, but highlighted the difficulties of obtaining accurate and relevant reports of this kind. Although 2 of the 26 owners initially selected for interview were not interviewed because they reported during the consultation that they had not observed any relevant behaviour changes, it is hypothesised that if questioned carefully some behaviour changes would have been

reported by these owners. Similarly, developers of the NCCPC (Breau *et al.*, 2000) found that many caregivers felt unable to describe how they detected pain yet quickly recognised relevant behaviours when offered a list of these from which to select. A similar effect was noted in this study when owners tended to select many more descriptors from a list than they were able to suggest without such a prompt. It was therefore considered essential to devise an instrument that would ask appropriate questions of dog owners in order to facilitate their accurate reporting of relevant behavioural observations. Once the dog owner was identified as the potential respondent for a proxy instrument to measure a dog's chronic pain, a key requirement of the instrument was that it should be able easily to be completed by any dog owner. It was therefore essential that instrument items would readily be understood by this group, and this was ensured by using dog owners to generate potential items for inclusion in the instrument.

In the preliminary interviews with dog owners, many of the behaviour changes were observed and reported by owners as changes in behavioural style, and the semi-structured interviews with owners confirmed this finding. Owners chose to use terms and phrases that described the attitude and demeanour of their dogs – styles of behaviour interpreted by owners as indicative of their dogs' mental states. It has been proposed that the most important element of an animal's experience of pain, in terms of welfare, is the significance of that pain to the individual (Rutherford, 2002). The suffering associated with pain is caused by its affective dimension – how it makes the animal feel. It was interesting, therefore, that the owners interviewed chose to interpret many of their behavioural observations as indicative of how their dogs were feeling. Of course, owners may have been mistaken in these interpretations, but if such interpretations were to prove reliable, then they would offer the opportunity to access the subjectivity of the sufferer, which is the goal of both pain and HRQL measurement.

Nearly 20 years ago, a working party of the Association of Veterinary Teachers and Research Workers (AVTRW) that was tasked with preparing guidelines for the assessment of animal pain, included in its recommendations that an agreed glossary of words should be prepared to describe a range of animal responses to pain, such as vocalisation, posture, gait and also those to describe the animal's 'mental state' (Association of Veterinary Teachers and Research Workers, 1986). In spite of this recommendation, no such comprehensive list of words to describe an animal's mental state has so far been devised. Yet, given the importance of mental state to the experience of suffering of any kind, in any species, such a glossary of words with shared meaning for that species may have the potential to be a useful resource in veterinary or human medicine.

The importance of the mental state of the sufferer has been recognised in human pain measurement, with tools such as the McGill Pain Questionnaire (Melzack, 1975) being the first to recognise the contribution of the affective dimension to the individual's experience of pain, and with the emergence of HRQL instruments, with their emphasis on the psychological and social well-being of the individual, as valid clinical measures of chronic pain. However, many tools designed for proxy use, such as those introduced on page 72 (RMH-POQLQ, Miami Pediatric Quality of Life Questionnaire, parent version of the PedsQL™, DEGR® Scale, PPPM, NCCPC), continue to focus to a large extent upon the reporting of overt behaviours rather than encouraging the respondent to interpret behaviour as an expression of mental state. In other words, many instruments ask about what the individual does, rather than about the style in which he or she does it, when the latter report may be more richly informative (Bateson, 1991; Martin and Bateson, 1993; Wemelsfelder, 1997; Wemelsfelder, 1999; Wemelsfelder, 2001a). Such limited use of human skills of interpretation of behaviour is perhaps surprising, particularly when the proxy is the parent, since it is likely that evolutionary pressures will have favoured reliable and sensitive mechanisms of interpreting subjective experiences in non-verbal offspring (Preston and de Waal, 2002).

Items for the proxy human pain and HRQL instruments listed above were obtained from those who had expert relevant knowledge, including clinicians and professional and lay carers. For example, items for the NCCPC were generated through semi-structured interviews with caregivers; those for the PPMP were generated through information gathering from parents using pain diaries; and interviews with nursing staff generated the items for the DEGR®. The items for the RMH-POQLQ were devised by the developers on the basis of pre-existing measures, but the relevance and adequacy of the items were subsequently checked by administering the prototype questionnaire to a sample of the target population of parents or caregivers of children with cancer, and further validated by pediatricians/pediatric oncologists and child psychologists. The items for the Miami Pediatric Quality of Life Questionnaire were generated through extensive videotaped interviews of families of children with cancer. The items for the PedsQL™ were generated by a search of the literature, open-ended interviews with patients and their families, and discussions with paediatric healthcare professionals. Expert panel discussion, involving neonatologists, nurses, psychologists and physiotherapists was used to create the items for the EDIN, developed for the clinical assessment by nursing staff of prolonged pain in premature infants (Debillon *et al.*, 2001).

The veterinary practitioners and owners whose participation in this study is described in this chapter were considered to be the most appropriate informants for the purposes of identifying behaviours relevant to the measurement of a dog's chronic pain, and of generating potential items for inclusion in an instrument for this purpose. Consequently, their role was considered to provide some measure of validity for the instrument developed from this information.

In reports of the development of the above proxy human pain and HRQL instruments, no indication was given of the extent to which those who supplied the instrument items were directed towards the kinds of

observations sought by the instrument developers, and it is possible that information was sought in such a way that a qualitative interpretation of behaviour was discouraged. It has been suggested (Theunissen *et al.*, 1998) that proxy instrument developers 'need to study exactly what [proxy raters]… do observe'. An author reviewing the literature on proxy evaluation of QoL (Cohen, 1999) concluded that manifestations of positive and negative affect in patients with dementia offer 'a window on the patient's subjective state' and may therefore be important in making proxy assessments of QoL for such individuals. The observation was made in this study that owners demonstrated a readiness to describe their dogs behaviour using terms that indicated that they were interpreting behaviour as manifestations of affect – as 'the expression of the emotions' (Darwin, 1872).

Because owners tended to report their observations in a way that described what they interpreted as the affective experience of their dogs, and because that experience is likely to be the most relevant dimension of pain as far as the sufferer is concerned, it was decided to generate a collection of the terms that owners used to make these kinds of report. The questionnaires DGQ1 and DGQ2 were designed to generate such a collection of terms: terms that dog owners used to describe the attitude and demeanour of their dogs when well and when unwell, and particularly when suffering chronic pain. DGQ2 asked owners to suggest and select words that they would use to describe their dogs if in chronic pain which, in spite of the risk of biased responses, was considered an important means of increasing the validity of the item generation process. The item pool thus generated was considered to represent a communal lexicon, the most appropriate collection of terms upon which to base a questionnaire (Clark, 1998; Garrod, personal communication).

An important consideration when developing an instrument for clinical use is utility. For that reason, it is valuable to base items on simple terms that will be readily understood, an approach taken with the EDIN (Debillon *et al.*, 2001). One difficulty of this approach is that the vocabulary with which people

familiarly describe emotion is invariably anthropomorphic. The descriptors owners chose to describe their dogs could, for the most part, equally appropriately be applied to a person suffering chronic pain. An anthropomorphic approach to assessing the subjective experiences of animals has been criticised in the past, but more recently it has been argued that an anthropomorphic approach can be helpful in formulating hypotheses that can subsequently be tested (see Chapter 1, pages 29 and 40). It has also been argued that anthropomorphism is a useful device that facilitates an intelligible discussion of the behaviour and emotions of animals (Bekoff, 2002), and it has been suggested that scientists often resort to a qualitative interpretation in order adequately to describe their quantitative results (Wemelsfelder and Farish, 2004).

An analysis of the responses obtained with the exploratory questionnaire described in this chapter revealed, by close examination of the alternative descriptions suggested by owners, and of additional information offered by 11 of the 17 owners, the following faults in questionnaire design: respondents were unable to select from descriptions of behaviour offered because none was suitable; respondents may have had difficulty choosing between descriptions because differentiation between descriptions of behaviour was poor; the inclusion within one item of more than one kind of behaviour (e.g. sociability and clinginess) may have made it difficult for respondents to answer; inaccuracy of recall of respondents over a period of several weeks between questionnaires may have made accurate responses difficult.

A component of the questionnaire was either a page on which owners were invited to suggest single words to describe their dogs' 'general attitude' over the previous week, or a page bearing the 'attitude and demeanour' list also incorporated in DGQ2. Of the 9 owners whose questionnaires included the former page, all made some suggestions (from 3 to 11 suggestions were made on these questionnaires). Of the 14 owners who were offered the 'attitude and demeanour' list, all were able to make selections from the list (from 6 to 38

selections were made on these questionnaires). It should be noted that some owners completed questionnaires of both types in the course of their dogs' treatment. Owners selecting from the 'attitude and demeanour' list did not appear to choose more words from the beginning of the list than from the end of the list (which might have indicated decreasing attention towards the end of the list). A number of owners chose to give particular emphasis to some of the descriptors they selected, e.g. by underlining those descriptors several times, as though in an effort to give particular weight to the applicability of those descriptors.

A review of the literature on chronic pain assessment in animals, along with the results of interviews with dog owners and veterinary practitioners, made the dog owner, familiar with the individual dog and able to observe the dog's behaviour in its usual environment, a good candidate for reporting on the widespread and often subtle behavioural disturbances associated with chronic pain. In interviews, descriptor-generating questionnaires, and the exploratory format questionnaire, owners had demonstrated their readiness to report their dogs' behaviour using terms that described mental state (for which the description 'subjective-expressive' terms is proposed) and some owners sought to apply a rating to such terms by combining them with adverbs, prefixes and suffixes which expressed relation or degree, or by graphic emphasis of some of their choices from a list of such terms.

These observations, combined with the difficulties revealed by the exploratory format questionnaire (of providing descriptions of behaviour that were applicable and adequately differentiated), led to a decision being made to construct an instrument based on the description of all relevant behaviours by means of rating the applicability of simple descriptive terms that included appropriate subjective-expressive terms. A collection of such terms to describe dogs in chronic pain and dogs that were healthy had been made, and formed an item pool from which the instrument would be developed.

In the case of this study, all qualitative interpretations of interview transcripts were carried out by the author. Although great care was taken over these interpretations, the potential for bias is recognised. However, the findings of the qualitative interpretations were fundamental to the development of the instrument, and so they would be validated by subsequent validation of the instrument itself.

Chapter 3

CONSTRUCTING THE INSTRUMENT

3.1 Introduction

The processes necessary for the creation of a health assessment instrument were outlined in Chapter 1 (pages 11-13), namely: specify measurement goals for the chosen patient population; identify relevant domains of measurement and create a pool of potential items; select suitable items for inclusion in the instrument; format the instrument (including selection of response options and details of administration); pre-test the instrument; field-test the instrument to establish its psychometric properties. Such steps were taken in the generation of recent proxy human pain and HRQL instruments such as the EDIN (Debillon *et al.*, 2001), the NCCPC (Breau *et al.*, 2000; Breau *et al.*, 2002a), the Royal Marsden Hospital Paediatric Oncology Quality of Life Questionnaire (Watson *et al.*, 1999) and the PedsQL™ (Varni *et al.*, 1999; Varni *et al.*, 2001).

Having generated a pool of potential items for an instrument to measure chronic pain in dogs suffering from chronic and painful conditions, as described in Chapter 2, the next steps were to select items from that pool, to validate that selection, and to construct a prototype instrument. This chapter describes these steps, beginning with an introduction to the range of considerations involved in developing an instrument with adequate psychometric properties.

The most important attribute sought by developers of a new instrument is validity – evidence that the instrument is able to measure what it was designed to measure. Various kinds of validity were introduced in Chapter 1 (pages 14-15). The content validity of an instrument is a measure of the appropriateness

of the collection and selection of its items (its content) and is therefore determined during instrument development (Coste, 1995). The content validity of a new instrument can be evaluated by an expert group who are asked to assess the relevance and adequacy of the items selected for inclusion in the instrument and to suggest any additions (or deletions) they deem to be necessary (Streiner and Norman, 1995).

When measuring continuous variables, such as pain and HRQL, there are several established ways in which questions can be designed to facilitate sensitive and accurate responses. The most frequently used direct estimation methods include NRS, VAS, adjectival scales (with or without a VAS) and Likert scales. On a Likert scale, responses are framed on an agree-disagree continuum. Traditionally, response options would be verbal, e.g. 'strongly disagree'/'disagree'/'no opinion'/'agree'/'strongly agree'. Sometimes these options are represented numerically, e.g. as they are in the parent form of the PedsQL™ (Varni *et al.*, 1999). Likert scales are commonly used in psychology and health measurement, their popularity due to their simplicity and track record in empirical studies (Ware, 1995).

It is recommended that the title of an assessment instrument should be concise and clear but should avoid revealing the project's hypothesis, in order to minimise bias (Vaillancourt *et al.,* 1991). It is essential to be aware of problems of readability, and the dangers of ambiguity in question wording, double-barrelled questions, jargon, value-laden words, and positive or negative wording (Payne, 1951; Streiner and Norman, 1995), and questions should aim to be comprehensible to someone with a reading age of 12 years (Streiner and Norman, 1995).

The inclusion of some transition items, where respondents are required to provide an estimate of change from one occasion to another, is recommended for human health measurement scales (Ziebland, 1994; Liang, 2000).

Once a best attempt has been made to construct an instrument with all of these considerations in mind, a key step is to pre-test the instrument to ensure that it is suitable for its intended respondents (Vaillancourt *et al.*, 1991). Various methods of pre-testing have proved useful as a check on whether or not questions have been fully understood, including asking respondents to rephrase questions in their own words, asking respondents to think aloud as they answer the questions, or asking respondents to complete the questionnaire and then to explain their responses to selected questions (Streiner and Norman, 1995). It is important that instruments are developed using subjects who are representative of its intended respondents. Selection biases should be avoided in the selection of subjects, and the consequences in this regard of non-responses, subjects lost to follow-up and missing data must be addressed.

It is important, when developing an instrument for clinical use, that it should have utility (as described in Chapter 1, pages 17-18), and the time taken to complete an instrument must be considered as part of such an assessment. This is reported for some proxy human instruments, e.g. the proxy version of the PedsQL™4.0 is reported as taking approximately 10 minutes to complete (Varni *et al.*, 2002a) and the DEGR® taking 5-10 minutes (Gauvain-Piquard *et al.*, 1999).

The question wording difficulties revealed in the exploratory format questionnaire that were described in Chapter 2 (pages 67, 70, 78), and the apparent readiness of owners to use a range of simple terms, many of which were subjective-expressive terms, to describe their dogs' behaviour, directed attention to the rating of such terms being a potentially valuable approach to the assessment of canine chronic pain using owner observations. It was decided to construct an instrument based on the description of all relevant domains of behaviour using such terms and the rating of these.

3.2 Methods and results

3.2.1. Creation of a matrix of behavioural domains and descriptors as a basis for the instrument

3.2.1.1Materials and methods

The 32 types of behaviour in which disturbances were identified as relevant to measuring chronic pain in dogs (Figure 2.1, page 56), were grouped by the author to form behavioural domains. In order to sample each of the domains thus created, using (as far as possible) descriptive terms that owners used most readily to describe behaviour expressive of their dogs' subjective experience, suitable descriptors were chosen from the pool of potential items generated using DGQ2 (presented in Appendix 7). The criteria used to select these descriptors were validated subsequently. The descriptors selected were those that owners used most readily to describe a dog in chronic pain (negative descriptors) or when well (positive descriptors), according to the following criteria: negative descriptors were those terms selected in DGQ2 by more than 33% of group A (see page 87 for allocation of owners to group A or B), along with additional terms selected by more than 33% of group B or suggested by more than one owner in the combined groups; positive descriptors were those terms not selected in DGQ2 by any owners in group A, and those selected by fewer than 10% of all owners in groups A and B.

Thereafter, consideration was given to whether or not any additional descriptors were required in order to describe all behaviours in terms of positive and negative descriptors, and to ensure that all relevant behaviours were adequately represented in the instrument.

3.2.1.2 Results

The 32 types of behaviour in which owners reported change were incorporated into 11 behavioural domains, namely, Activity, Comfort, Appetite, Extroversion/Introversion, Aggression, Anxiety, Alertness, Dependence, Contentment, Consistency and Agitation (Table 3.1).

Table 3.1 Domains of behaviour that were proposed to encompass the behaviour changes of 17 dogs diagnosed as suffering from chronic degenerative joint disease, when changes were identified through semi-structured interviews with their owners (n=17).

Proposed behavioural domain	**Behaviour changes**
Activity	Activity, mobility, agility, daytime sleeping/resting, stamina, playfulness, keenness to exercise, excitability, pattern of daytime activity, scent marking, grooming
Comfort	Pain-related vocalising
Appetite	Drinking, appetite
Extroversion/introversion	Sociability towards family members, sociability towards people outside the family, interactions with other animals
Aggression	Aggression
Anxiety	Anxiety/fearfulness, compulsive behaviour
Alertness	Curiosity, obedience
Dependence	Attention-seeking behaviour, 'clinginess'
Contentment	Attitude and demeanour, facial expression, posture
Consistency	Inconsistency of behaviour
Agitation	Restlessness, panting, night-time sleeping

A matrix was formed by allocating, to each of these domains, selected negative and positive descriptors that had been generated using DGQ2. Applying the criteria established for selection of descriptors, described above (page 84), 54 negative descriptors were considered for inclusion in the matrix and are shown in Table 3.2. Of these negative terms noted, it was decided not to include 'lazy' in the matrix, since 'lethargic' was considered by the author to describe this demeanour without suggesting any underlying personality trait. 'Dependent' and 'lies curled in' were also not included in the matrix, since the author considered that these terms could be used with no intention to indicate behavioural style, and there were considered to be sufficient similar items. 'Pained' was included rather than 'painful', since 'painful' could refer to the cause of the pain and not to the behaviour of the dog. The final selection of negative descriptors from those generated using DGQ2 is shown in Table 3.3. Of those descriptors, 59% were also suggested by dog owners completing descriptor-generating questionnaire DGQ1 as terms they would use to describe the attitude and demeanour of their dog when it was unwell, and these descriptors are emboldened in Table 3.3.

'Compulsive' was not suggested by any owner and was not offered in the 'attitude and demeanour' list generated by DGQ1, but was identified by an experienced pet behaviour therapist as a common behaviour disturbance of a dog with chronic pain (Lindley, personal communication), and so this term was identified for inclusion in a matrix of behavioural domains and descriptors relevant to measuring chronic pain in dogs.

In order to select positive descriptors for inclusion in the matrix, the selection criteria previously described (page 84) were applied to all descriptors generated using descriptor-generating questionnaire DGQ2. As a result, 37 terms were considered for inclusion as positive descriptors in the matrix, and these are shown in Table 3.4.

Table 3.2 Descriptors generated using the descriptor-generating questionnaire DGQ2 and meeting the criteria for inclusion as negative descriptors in a matrix of behavioural domains and descriptors relevant to measuring chronic pain in dogs.

Criterion for selection	Descriptors selected
Terms selected in DGQ2 by >33% of owners in group A	***Agitated, anxious, attention-seeking, cautious, clingy, dependent, depressed, dull, irritable, lacklustre, lethargic, listless, quiet, nervous, sad,*** **strained,** ***subdued, tired, uneasy, unhappy, uninterested, unsettled, withdrawn***
Additional terms selected in DGQ2 by >33% of owners in group B	**Panicky, restless sorrowful, unsociable, weary**
Additional terms suggested in DGQ2 by >1 owner in the combined groups	**Accepting, aggressive, apathetic, comfort-seeking, confused, crying, detached, distressed, frightened, grumpy, lazy, lies tucked in/curled up, miserable, moaning, pained/painful, panting, pathetic/pitiful, sleepy, slowed, sluggish, stoical, stubborn, off his/her food, unresponsive, upset, whining**

Bold indicates a term suggested by more than one owner, or selected by more than 33% of group A or B, ***bold italic*** indicates a term selected by more than 33% of both groups, and underlined means selected by more than 33% of one or other or both groups, and also suggested by more than one owner.

Group A (n=14): owners who reported that their dogs were at the time of completing DGQ2, or had previously been, suffering from a condition considered by the author to be chronic and likely to be painful.

Group B (n=53): owners who may or may not have had direct experience of a dog suffering chronic pain.

Table 3.3 Negative descriptors selected for inclusion in a matrix of behavioural domains and descriptors relevant to measuring chronic pain in dogs.

Emboldened descriptors are those also appearing in the UNWELL list from DGQ1

Accepting, **aggressive, agitated, anxious,** apathetic, **attention-seeking**, cautious, **clingy,** comfort-seeking, confused, **crying, dependent, depressed,** detached, distressed, **dull, frightened,** grumpy, **irritable, lacklustre, lazy, lethargic,** lies tucked in/curled up, **listless,** miserable, moaning, **nervous,** off his/her food, pained/painful, **panicky,** panting, pathetic/pitiful, **quiet, restless, sad, sleepy, slowed,** sluggish, **sorrowful**, stoical, **strained,** stubborn, **subdued, tired, uneasy, unhappy, uninterested,** unresponsive, **unsettled,** unsociable, upset, **weary,** whining, **withdrawn**

Table 3.4 Descriptors generated using the descriptor-generating questionnaire DGQ2 and meeting the criteria for inclusion as positive descriptors in a matrix of behavioural domains and descriptors relevant to measuring chronic pain in dogs.

Criterion for selection	Descriptors selected
Terms selected in DGQ2 by no owners in group A	Active, alert, boisterous, bold, bouncy, bright, comical, consistent, contented, eager, energetic, excitable, fun-loving, greedy, happy, independent, inquisitive, interested, keen, lively, nosy, playful, sociable, unpredictable
Terms selected in DGQ2 by <10% of owners in groups A and B	Affectionate, calm, cheeky, curious, friendly, good-natured, laid-back, loving, mischievous, noisy, obedient, outgoing, placid, relaxed

Group A (n=14): owners who reported that their dogs were at the time of completing DGQ2, or had previously been, suffering from a condition considered by the author to be chronic and likely to be painful.

Group B (n=53): owners who may or may not have had direct experience of a dog suffering chronic pain.

Of those terms considered, 'mischievous', 'cheeky', 'loving' and 'noisy', which were selected by fewer than 10% of owners in groups A and B, and 'comical', which was selected by no owners in group A, were not chosen for inclusion as positive descriptors to be included in the matrix since the author considered them to be difficult to allocate to a behavioural domain or because there appeared already to be sufficient descriptive terms to adequately describe the domain in which they would be included. 'Unpredictable' was also selected by no owners in group A but was selected by 17% of group B and had appeared in the UNWELL list from DGQ1 and was therefore considered not to be a positive descriptor. The final selection of positive descriptors is shown in Table 3.5. Of those descriptors, all but 1 were also suggested by dog owners completing descriptor generating questionnaire DGQ1 as terms they would use to describe the attitude and demeanour of their dog when it was well, and these descriptors are emboldened in Table 3.5.

These selected positive descriptors, along with the negative descriptors already identified, were allocated to the proposed domains of behaviour in order to form a matrix of behavioural domains and descriptors relevant to measuring chronic pain in dogs. In order to ensure that each of the domains was comprehensively covered by both positive and negative descriptors, some additional words were included by the author, chosen as far as possible from amongst all words previously suggested by owners. Thirteen additional descriptors (all to be subsequently validated) were added to balance all domains in terms of positive and negative descriptors and to ensure that all relevant domains were adequately represented in the instrument. These additional descriptors, chosen as far as possible from amongst the words that had been suggested by owners in questionnaires or interviews (emboldened, below), were 'comfortable', 'complaining' 'compulsive', '**confident**', '**easy-going**', 'even-tempered', interested in food', '**picky** (food)', '**sore**', '**stiff**', 'tireless', '**uncomfortable**' and '**unpredictable**'; others were proposed by the author.

Table 3.5 Positive descriptors selected for inclusion in a matrix of domains and behavioural descriptors relevant to measuring chronic pain in dogs.

Emboldened descriptors are those also appearing in the WELL list from DGQ1.

Active, affectionate, alert, boisterous, bold, bouncy, bright, calm, consistent, contented, curious, eager, energetic, excitable, friendly, fun-loving, good-natured, greedy, happy, independent, inquisitive, interested, keen, laid-back, lively, nosy, obedient, outgoing, placid, playful, relaxed, **sociable**

Finally, although 'accepting' was a term owners had used to describe a dog with chronic pain, it was felt that an owner who used this descriptor was suggesting that the dog was coping well with its circumstances, suggesting less suffering than a dog that was not so described. It was therefore decided to include this term as a positive descriptor, since a dog considered to be 'accepting' would suggest less suffering, in common with other positive descriptors.

A total of 96 descriptors were thus included in the matrix within the 11 proposed behavioural domains. This matrix is shown in Table 3.6.

3.2.2 Expert validation of the matrix

3.2.2.1 Methods

To assess the relevance and adequacy of these potential items for inclusion in an instrument to measure canine chronic pain, the matrix of behavioural domains and descriptors was subjected to validation by 7 veterinary general practitioners, 5 veterinary specialists, and by 10 owners with experience of dogs with chronic and painful conditions. These owners were recruited by 5 veterinary general practitioners if their dogs met the criteria: 'have recently suffered or are currently suffering chronic pain'. The members of this expert group were asked to assess whether any behavioural domains were missing from the matrix or whether any of the included domains were irrelevant. They were also asked to suggest, for each behavioural domain, any additional descriptors they considered necessary to describe it, and to comment on descriptors they considered irrelevant. Finally, they were asked to comment on whether any descriptor had been included in an inappropriate domain.

3.2.2.2 Results

Of the expert group asked to validate this matrix, two veterinary practitioners and one owner considered that the matrix required no revision. The other 19 validators individually suggested few changes, but there was no consensus among them.

Table 3.6 Matrix of behavioural domains and associated descriptors relevant to measuring chronic pain in dogs, generated using descriptor-generating questionnaire DGQ2 or suggested by an experienced pet behaviour therapist, prior to validation by an expert group consisting of veterinary general practitioners and specialists, and owners of dogs with experience of chronic pain.

NOTE: Bold indicates a term suggested by more than one owner, or selected by more than 33% of group A or B, ***bold italic*** indicates a term selected by more than 33% of both groups, and <u>underlined</u> means selected by more than 33% of one or other or both groups, and also suggested by more than one owner. *Italic* indicates a term not selected by any owners in group A or by fewer than 10% of all owners in groups A and B. Other terms added by author.

Behavioural domains	Negative descriptors	Positive descriptors
Activity	**Apathetic** ***<u>Lacklustre</u>*** ***Lethargic*** ***<u>Listless</u>*** **Sleepy** **Slowed** **Sluggish** ***<u>Tired</u>*** **Weary**	*Active* *Boisterous* *Bouncy* *Energetic* *Lively* *Playful* Tireless
Comfort	Complaining **Moaning** **Pained** Sore Stiff **Stoical** Uncomfortable	Comfortable
Appetite	**Off his/her food** Picky	*Greedy* Interested in food
Extroversion/introversion	**Detached** ***<u>Quiet</u>*** ***<u>Subdued</u>*** **Unsociable** **Unresponsive** ***<u>Withdrawn</u>***	*Affectionate* *Bold* *Curious* *Eager* *Excitable* *Friendly* *Fun-loving* *Nosy* *Outgoing* *Sociable*

Table 3.6 Continued from page 93.

Behavioural domains	Negative descriptors	Positive descriptors
Aggression	**Aggressive** **Grumpy** ***Irritable*** **Stubborn**	Even-tempered *Good-natured* *Placid*
Anxiety	***Anxious*** ***Cautious*** **Frightened** ***Nervous*** **Strained** ***Uneasy***	**Accepting** Confident Easy-going *Laid-back*
Alertness	**Confused** ***Depressed*** ***Dull*** ***Uninterested***	*Alert* *Bright* *Inquisitive* *Interested* *Keen* *Obedient*
Dependence	***Attention-seeking*** ***Clingy*** **Comfort-seeking** **Pathetic/pitiful**	*Independent*
Contentment	**Miserable** ***Sad*** **Sorrowful** ***Unhappy***	*Contented* *Happy*
Consistency	Unpredictable	*Consistent*
Agitation	***Agitated*** **Compulsive** **Crying** **Distressed** **Panicky** **Panting** **Restless** ***Unsettled*** **Upset** **Whining**	*Calm* *Relaxed*

All comments received were discussed in a focus group consisting of the author and her three supervisors who are experienced in animal pain and welfare assessment. Consequently, a number of revisions to the matrix (both domains and descriptors) were agreed, as described below.

In response to comments received from vets, 'stubborn' was removed and 6 new descriptors were added: 'reluctant', 'disturbed', 'awkward', 'athletic', 'fit' and 'at ease'. Also 'unpredictable' was replaced by 'inconsistent', as the directly opposite term to 'consistent'. In response to comments received from owners, no terms were removed (since the comments of most owners were based on limited experience of dogs suffering chronic pain) and the following 6 new terms/phrases were added: 'apprehensive', 'groaning', 'thirsty', 'enthusiastic about food', 'resigned', and 'limping'.

In response to comments from vets, a Compulsion domain and a Posture/Mobility domain were added, and some descriptors were moved from one domain to another, ('stiff' and 'relaxed' were moved to Posture/Mobility; 'panicky', 'distressed' and 'upset' to Anxiety; 'confident' to Dependence).

At this stage, two further descriptors were added by the author: 'territorial/protective' to Aggression (to allow for a positive form of aggression), and 'stretching' to Comfort, since this had been a behaviour observed by a veterinary acupuncture specialist to change with relief from chronic pain (Hutchison, personal communication).

These revisions resulted in the revised matrix shown in Table 3.7, containing a total of 109 descriptors, distributed between 13 behavioural domains.

Table 3.7 Matrix of behavioural domains and associated descriptors relevant to measuring chronic pain in dogs, following expert validation by 7 veterinary general practitioners, 5 veterinary specialists and 10 owners of dogs with experience of chronic pain.

Behavioural domains	**Negative descriptors**	**Positive descriptors**
Activity	Apathetic Apprehensive Lacklustre Lethargic Listless Reluctant Sleepy Slowed Sluggish Tired Weary	Active Boisterous Bouncy Energetic Lively Playful Tireless
Comfort	Complaining Groaning Moaning Pained Sore Stoical Uncomfortable	Comfortable Stretching
Appetite	Off his/her food Picky (food)	Greedy Enthusiastic about food Interested in food Thirsty
Extroversion/introversion	Detached Quiet Subdued Unsociable Unresponsive Withdrawn	Affectionate Bold Curious Eager Excitable Friendly Fun-loving Nosy Outgoing Sociable
Aggression	Aggressive Grumpy Irritable	Even-tempered Good-natured Placid Territorial/protective

Table 3.7 Continued from page 96.

Behavioural domains	Negative descriptors	Positive descriptors
Anxiety	Anxious Cautious Distressed Frightened Nervous Panicky Strained Uneasy Upset	Accepting Easy-going Laid-back
Alertness	Confused Depressed Dull Uninterested	Alert Bright Inquisitive Interested Keen Obedient
Dependence	Attention-seeking Clingy Comfort-seeking Pathetic/pitiful	Confident Independent
Contentment	Miserable Resigned Sad Sorrowful Unhappy	Contented Happy
Consistency	Inconsistent	Consistent
Agitation	Agitated Crying Disturbed Panting Restless Unsettled Whining	Calm At ease
Posture/mobility	Awkward Limping Stiff	Athletic Fit Relaxed
Compulsion	Compulsive	

3.2.3 Assessment of readability of descriptors

3.2.3.1 Methods

Most (87%) of the descriptors ultimately included in the validated list (the 95 descriptors that were included from the pre-validation matrix) were tested for readability by reference to *Collins Junior Dictionary* (age range 9-12 years), *Oxford Children's Dictionary* (age range 9-11 years), and the *Longman Active Study Dictionary* (for students of English as a second language), and by consulting a group of (17) adult literacy tutors and a group of (12) mixed-ability 9-year old school pupils.

3.2.3.2 Results

The assessment of readability of descriptors was inconclusive. Although some (36) descriptors were absent from some dictionaries and some (10) were considered by one or other consulted group to cause reading difficulties, no words fell into all of these categories. Because of this lack of agreement, no terms were excluded on grounds of readability.

3.2.4 Construction of the prototype instrument – GUVQuest–Dog

3.2.4.1 Methods

A prototype owner questionnaire was designed, based on the validated list of descriptors. This prototype was pre-tested in UGSAH using a double-interview technique. Owners were selected to be involved in pre-testing by participating clinicians in orthopaedic, soft tissue and oncology clinics, who were asked to identify owners of dogs diagnosed with chronic and painful conditions. Each of 26 dog owners completed the questionnaire with the author present. Owners were encouraged to comment both while completing the questionnaire and once the questionnaire was completed. On completion, owners were asked to explain a selection of their answers. The design of the questionnaire was revised during pre-testing, and pre-testing ended when it seemed that an optimum design had been achieved.

3.2.4.2 Results

The prototype instrument took the form of a structured questionnaire based upon the validated list of descriptors, with each descriptor having an associated 7-point Likert-type scale, from 0 to 6, to allow the owner to rate how well each of the 109 descriptors described his or her dog. A score of 6 indicated '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). All of the descriptors were presented in one alphabetically ordered list.

In addition to questions designed to obtain demographic information, the prototype also included some transition questions. These asked the owner to rate any global change in a range of behavioural domains that were among those hypothesised to be affected by chronic pain and were considered by the author to be able readily to be understood as global domains by potential respondents: Activity, Pain, Sociability, Aggression, Anxiety, Enthusiasm, Happiness and Mobility. Owners were asked to respond to each of these questions using a 7-point verbal rating scale (greatly decreased; decreased; slightly decreased; no change; slightly increased; increased; greatly increased).

Demographic information about the dog, the owner and the family was also requested within the prototype instrument (GUVQuest–Dog), which was titled 'Glasgow University Health-Related Dog Behaviour Questionnaire'.

The design of the questionnaire was revised four times during pre-testing, and the end design incorporated a number of improvements on the prototype. Because the meanings of some of the descriptors changed with context (for example, one owner in pre-testing commented that 'reluctance depends on what you're asking her to do') it was decided to place each descriptor within an appropriate context as an aid to understanding. The items, originally arranged alphabetically, were now each placed into one of 13 familiar contexts, namely, Activity, Eating and drinking, Discomfort, Response to

owner, Extroversion, Irritability, Alertness, Calmness, Contentedness, Mobility, Anxiety, Consistency and Stoicism. While, in terms of the descriptors they contained, some of these contexts nearly or exactly matched the domains of behaviour previously identified as relevant to the measurement of chronic pain in the dog, the purpose of the contexts was simply to increase readability: they were not intended to represent those behavioural domains.

To be specific, the descriptors included in the *Eating and drinking* context included in the instrument matched exactly those contained in the 'Appetite' domain created by the author. In terms of the descriptors they contained, the *Irritability* context matched exactly the 'Aggression' domain, the *Alert* context matched exactly the 'Alertness' domain, and the *Contented* context matched exactly the 'Contentment' domain. However, descriptors contained within the 'Extroversion/introversion' and 'Dependence' domains were largely grouped, in the instrument, into two contexts: *How your dog behaved towards you* (response to owner) and how *Outgoing* (extroversion) was your dog. Such placing in context helped to clarify the meaning of terms like 'comfort-seeking', which were intended to be associated with owner interaction rather than physical comfort. All but three of the descriptors contained within the 'Extroversion/introversion' domain were included in the *Outgoing* context. Those three were, along with all but two of the descriptors contained within the 'Dependence' domain, included in the *How your dog behaved towards you* context. Of those two, 'pathetic/pitiful' was placed in the *Outgoing* context and 'confident' was placed in the *Anxious* context. The *Anxious* context included all the terms of the 'Anxiety' domain, with the addition of 'confident' from the 'Dependence' domain, and with the exception of 'accepting', which was included in a context that referred to a longer period over which the rating should be made.

In most cases, respondents were asked to rate their dog's behaviour as it had been during the previous day. However, for descriptors such as 'accepting' it

was felt that a longer rating period was required, and a two-week period was selected. Also included in this context, *How your dog has been over the past two weeks* (stoicism), were 'consistent' and 'inconsistent' from the 'Consistency' domain, and 'stoical', from the 'Comfort' domain.

The *Calm* context contained the 'Agitation' domain and, for convenience, included the single term ('compulsion') of the 'Compulsion' domain. The *How your dog looked when it was moving* context contained all of the descriptors from the 'Posture/mobility' domain along with the descriptors 'tired' and 'apprehensive' from the 'Activity' domain. Other than those two terms, all terms included in the *Activity* context matched those in the 'Activity' domain.

Additional changes were made to the prototype during pre-testing: changes were made to the wording of questions designed to obtain demographic information, in order to clarify meaning; changes were made to instructions, for clarification; and changes were made to layout to improve ease of use. A sample page from the finished design is shown in Figure 3.1.

A short clinician questionnaire was devised in order to obtain information to aid the selection of suitable cases for a longitudinal study, and to obtain clinicians' pain ratings (using a familiar 10-point NRS) and estimates of change over time (using a 7-point VRS), and to gather relevant information about dogs in a control group.

Four slightly different versions of the clinician and of the owner questionnaires were prepared, for use with dogs suffering chronic pain or with healthy dogs, and for use on an initial occasion or on subsequent occasions. The pages of the owner questionnaire bearing the descriptor rating questions were identical in all four versions. Differences in other questions are shown in Appendix 9. The questionnaires are included in Appendix 10.

Figure 3.1 Sample page from prototype instrument: GUVQuest–Dog.

The following questions are about **HOW YOUR DOG BEHAVED TOWARDS YOU** yesterday.
Please answer each question carefully. Circle the number that shows how well each word describes your dog **as he/she was yesterday**.
Remember: 0 always means 'not at all…' and 6 always means 'couldn't be more…'
If you are unsure of the meaning of any question **please draw a line through that whole question.**

not at all **affectionate**	0	1	2	3	4	5	6	couldn't be more **affectionate**
not at all **attention-seeking**	0	1	2	3	4	5	6	couldn't be more **attention-seeking**
not at all **clingy**	0	1	2	3	4	5	6	couldn't be more **clingy**
not at all **comfort-seeking**	0	1	2	3	4	5	6	couldn't be more **comfort-seeking**
not at all **detached**	0	1	2	3	4	5	6	couldn't be more **detached**
not at all **independent**	0	1	2	3	4	5	6	couldn't be more **independent**
not at all **unresponsive**	0	1	2	3	4	5	6	couldn't be more **unresponsive**

3.3 Discussion

The criteria for the initial selection of items for the validation matrix (and the behavioural domains to which they were allocated) were devised by the author, but the matrix of domains and descriptors thus created was then subjected to validation by an expert group, and the validation of items selected for inclusion in the instrument would be evaluated further through field-testing of the prototype instrument. A similar approach has been used in the development of human pain and HRQL instruments (Melzack, 1975; Juniper *et al.*, 1997; Armstrong *et al.*, 1999).

During the process of expert validation a number of changes were made to the matrix of behavioural domains and descriptors, including the creation of two new domains. These changes, particularly the addition of the domain of Posture/mobility, which encompassed 4 new descriptors, might be regarded as substantial. However, it was considered important not to exclude at an early stage any domains or descriptors that might contribute to the measurement of interest. It is an important element of the validity of an instrument that its items adequately cover all of the relevant domains. In order to maximise validity, ensuring adequate coverage of domains may entail accepting some degree of redundancy of items included in the instrument, with reliability of assessment of a domain likely to increase with larger numbers of items (Streiner, 1993). Thus the expansion of the matrix during expert validation was seen as a valuable development, in the knowledge that the instrument developed from its domains and descriptors would subsequently be required to demonstrate other kinds of validity.

The number of items required by an instrument comprehensively to sample all relevant domains of HRQL means that such human instruments are often time-consuming to complete, which can compromise their utility (Streiner, 1993). This problem has been recognised by developers of human HRQL instruments, whose strategy for improving utility has been to develop shorter forms of the instruments by selecting key items from the originals. However,

shortening an instrument in this way can lead to a loss of validity (Ware and Sherbourne, 1992; Coste *et al.*, 1995).

Items included in instruments to measure human HRQL and pain, with the exception of the McGill Pain Questionnaire (Melzack, 1975), usually take the form of questions or statements that vary in length and complexity, with associated response options. The longer and more complex the item, the more time-consuming it is to read and to respond to. Interestingly, some psychometric instruments, including the widely-used Hospital Anxiety and Depression scale (HAD) (Zigmond and Snaith, 1983) and the Insights evaluator (a psychometric instrument designed to measure personality) (Insights Learning and Development Ltd., 1992-2003) instruct respondents to answer quickly, rather than to think too carefully about their responses, since their initial thoughts will more likely be the correct ones. In this respect, a dependence on wordy or complex items that take longer to read and understand may hinder the process of accessing relevant respondent observations. More rapidly understood items may facilitate access to unconscious information, a potentially valid and rich source of respondents' perceptions (Cleermans 2001; Reber and Perrig, 2001). Basing all of the GUVQuest's items on simple terms rather than on more complex and lengthy questions meant that the items could be read and understood relatively quickly, offering the possibility of accessing unconscious information.

Because each item of the GUVQuest consisted of a simple, familiar term accompanied by a ubiquitous 7-point rating scale, the response to each item could be obtained relatively speedily. In the case of the GUVQuest, the entire instrument, including its 109 core items, could be completed within 30 minutes, and the pre-testing reported in this chapter demonstrated that this was acceptable to respondents. Consequently, for current users it did not appear to be necessary to shorten the GUVQuest for reasons of utility, so avoiding the potential validity problems associated with that process. However, it is recognised that the use of the GUVQuest with populations of

respondents other than dog owners visiting a referral hospital (for example, dog owners visiting veterinary general practitioner surgeries) may reveal utility problems that will have to be addressed.

The use of simple descriptive terms rather than wordier questions avoided most of the potential problems that must be addressed in question wording (ambiguity, double-barrelled questions, jargon, value-laden words, positive or negative wording) but the readability of the chosen terms remained an important consideration. If a questionnaire is to be used by the general public then questions should be capable of being read by those whose reading ability is at the lower end of the normal range. In the assessment of readability of terms in the pre-validation matrix, that contained most of the items that were subsequently included in the GUVQuest, it was found that not all terms appeared in all three dictionaries to which reference was made (two children's dictionaries and one dictionary for students of English as a second language). In addition, some terms were gauged as being difficult by the group of mixed-ability 9-year-old pupils or by a group of adult literacy tutors that were consulted, yet had readily been used by dog owners. Because of this lack of agreement it was decided not to exclude any of the selected items from the instrument on the grounds of poor readability. However, it remains a possibility that some of the descriptors included may not easily be read and may not properly be understood by all of those for whom the questionnaire is intended, so that the ratings applied to such descriptors will be unreliable and invalid. An analysis of responses to individual items will be required in order to identify any items with which respondents appeared to be having this kind of difficulty.

In designing an instrument, an important consideration is the kinds of response options to be provided for each item. If responses are likely to lie on a continuum (rather than being categorical), as is the case with continuous variables like pain and HRQL, it is important to provide the opportunity for respondents to answer in this way, by providing a continuum of answer

options. Studies have shown that providing insufficient answer options makes a questionnaire more difficult to complete, reduces the information it can provide, introduces error and reduces efficiency (Streiner and Norman, 1995). The simple and widely-used Likert-type scale, using numerically presented options, was considered to be appropriate for the GUVQuest, and a seven-point scale was chosen. There is evidence that offering around seven answer options tends to result in good reliability in scales in which people are asked to discriminate unidimensional stimuli or single attributes (Cichetti *et al.*, 1985; Preston and Coleman, 2000). Although a 7-point scale is not widely used in health measurement, one recent publication (McClain, 2002) has suggested that such a scale should be considered by those seeking to develop reliable pain measures for children.

There is a tendency for labelled points or boxes on a scale to be endorsed more frequently than unlabelled ones, and for end-anchored scales (scales where only the end boxes are labelled) to pull responses towards the ends (increasing variability), although there may also be an 'end aversion bias', where respondents tend to avoid the extremes of a scale (Streiner and Norman, 1995). A disadvantage of adjectival scales is that the descriptive words or phrases may mean different things to different people and within different contexts. However, if numbers are to be used to help respondents to select a point on a scale, research has shown that negative numbers tend to be avoided by respondents (skewing the results) and so should be avoided (Schwarz *et al.*, 1991). Finally, it has been suggested (Matza *et al.*, 2004) that it may be good practice to reverse the order of scale responses for some questions in order to detect respondents who are not taking care to answer correctly and are simply choosing similar responses each time. The GUVQuest's association with each descriptor, positive or negative, of a simple 7-point numerical scale, from 0 to 6, addresses the range of considerations just outlined.

For an instrument to have face validity its items must appear on the surface to be measuring what they really are measuring. The advantage of this kind of validity is that it increases the acceptance of the instrument by the user, who can see easily the relevance of each item. A disadvantage of face validity, and also of direct estimation scales, is that when the intention of questions is obvious, the risk of respondent bias is higher (Streiner and Norman, 1995), which can increase instrument reliability while decreasing validity (Dijkers, 1999). The problem of respondent bias is widely recognised in health and social science (Sprangers and Aaronson, 1992; Sandvik *et al.*, 1993; Gibson *et al.*, 1999; Rogler *et al.*, 2001), has been noted in those required to quantify animal pain (Peterson, 2004), and has been identified as one that must not be neglected by instrument developers (Gibson *et al.*, 1999; Rogler *et al.,* 2001). However, steps can be taken in instrument design to make it more difficult to respond in a consistently biased manner even when individual items have face validity, for example by including a large number of items in the instrument, as is the case for the GUVQuest with its 109 core items. The inclusion of both positive and negative items for most domains, and the reversal of the meaning of the Likert-type scale depending on whether the item is a positive or a negative descriptor, should also make responding in a consistently biased fashion more difficult, as well as making it possible to detect those who are not answering carefully. Finally, in order to limit the risk of response bias, as recommended by Vaillancourt and colleagues (1991), the title of the prototype GUVQuest – 'Glasgow University Health-related Dog Behaviour Questionnaire' – made no mention of pain or of assessment, since it was felt that these concepts might bias respondents, either consciously or unconsciously.

The descriptive terms identified as potential items for the assessment instrument were not defined, since all have a dictionary definition and are in everyday use. Unlike existing human pain and HRQL instruments (with the exception of the MPQ), each item in the GUVQuest principally consists of a simple descriptive term, rather than a longer description. Because the meaning

of some terms is context-dependent, within the GUVQuest each term was associated with a simple, familiar context. The difficulty of comprehending words when they are out of context has been recognised by Clark and colleagues (1995) in connection with the MPQ, which is similarly based upon simple descriptive terms. Some of the contexts created for the GUVQuest did match domains of behaviour proposed in the matrix shown in Table 3.7, but contexts were not chosen for this reason but simply for their familiarity and to maximise understanding.

The next step in instrument construction was field-testing, to establish further the validity of the GUVQuest, and to explore its sensitivity and reliability as a clinical tool.

Chapter 4

FIELD-TESTING OF INSTRUMENT, AND ASSESSMENT OF VALIDITY OF INCLUDED MEASURES OF CLINICAL STATUS AND CLINICAL CHANGE

4.1 Introduction

There are various methods of evaluating the validity of a new instrument. An evaluation of the content validity of the GUVQuest was described Chapter 3. Construct validity, which was introduced in Chapter 1 (page 15), is also sought by instrument developers. The demonstration of construct validity requires that the scores obtained with an instrument should reveal predictable profiles and patterns of change that confirm the hypothesis upon which the instrument was constructed, so that the testing of these predictions reveals the extent to which the instrument does appear to be measuring that which it was intended to measure. Consequently, construct validity can only be explored through the field-testing of a new instrument.

For example, the purpose of the GUVQuest was to measure chronic pain, through its impact upon a range of domains of canine HRQL, through proxy rating of relevant behaviours by the dog's owner. The most important function of the tool would be to detect relevant changes in an individual dog over time, using the dog as its own control. It was hypothesised that the output from the instrument would be in the form of a profile of HRQL domain scores that would be applicable to that combination of dog and owner, and that changes in these scores would be sensitive to clinical change in a dog's chronic and painful condition. Testing an instrument's ability to detect change over time within a patient requires a longitudinal study. Such a longitudinal study was devised to field-test the GUVQuest.

Field-testing must be carried out with a population and in a manner that mirrors the use to which the validated instrument would be put in a clinical setting. Because it was intended that the instrument would be used in specialist clinics and in veterinary general practices, and because it should have some validity for measuring chronic pain of any cause, it was considered important to include in the field-testing dogs with chronic pain of a variety of causes, presenting in different clinical settings and prescribed a range of treatment options.

The core of the GUVQuest consisted of the 109 descriptor items and associated scales that were to be used by the owner to report observations of a dog's behaviour. Instrument scores would be calculated using the data generated by this part of the questionnaire, and these scores would be used to assess the construct validity of the instrument, by comparing scores for healthy dogs with scores for dogs diagnosed with a chronic and painful condition, and by comparing scores for an individual dog over time.

In order to provide independent measures of clinical status and clinical change with which to compare scores obtained with the core instrument items, a number of additional questions were included in the instrument. These questions, incorporated in the owner's questionnaire and in a brief clinician's questionnaire completed on the same occasion, were the clinician's pain score, the clinician's assessment of change, the owner's rating of pain/no pain, and the owner's transition questions.

The analysis of the data obtained from these additional questions was designed to validate these independent measures, by evaluating the following hypotheses:

Hypothesis 1

It was hypothesised that clinician pain scores for dogs receiving treatment for a chronic and painful condition would tend to decrease while treatment continued. This made the assumption that treatment was effective: while this

may not have been the case at the individual level it was likely to be so at the group level.

Hypothesis 2

It was hypothesised that there would be a relationship between a clinician's pain scores and that clinician's assessment of clinical change at each clinical examination.

Hypothesis 3

It was hypothesised that 'yes' responses to the owner's question 'do you think your dog is in any pain?' would be associated with higher clinician pain scores than would 'no' responses.

Hypothesis 4

It was hypothesised that owners' responses to transition questions regarding the behavioural domains of activity, pain, sociability, aggression, anxiety, enthusiasm, happiness and mobility would reflect relevant changes associated with the onset of a chronic and painful condition and with treatment of such a condition, and that these would be different from the responses of owners of healthy control dogs.

Hypothesis 5

It was hypothesised that there would be a degree of association between ratings awarded for most of the owner transition questions, since all of these behavioural domains were hypothesised to be affected by chronic pain and would be expected to co-vary with change in clinical condition.

Hypothesis 6

It was hypothesised that owner responses to transition questions would demonstrate some association with clinician assessments of clinical change, since changes in these behavioural domains were expected to be observed with clinical improvement or deterioration.

4.2 Field-testing

4.2.1 Materials and methods

4.2.1.1 Questionnaire formats

The GUVQuest existed in 4 forms. One version was designed for completion at the first consultation, while a modified version for completion at each follow-up consultation was also created. These were required for each of two types of respondent: owners of dogs receiving treatment for a chronic and painful condition, and owners of healthy dogs that were not suffering chronic pain.

The core element for each version of the questionnaire was identical, consisting of the 109 descriptor items and their associated rating scales. The additional questions included in each version differed in the following principal ways:

1) Initial owner questionnaires included demographic questions about the dog and its environment, and about the respondent, that were not repeated in follow-up questionnaires.

2) In the initial questionnaires for dogs receiving treatment for chronic and painful conditions, the owner's transition questions related to changes perceived since the dog became unwell. Initial questionnaires for healthy dogs did not include transition questions. The transition questions included in follow-up questionnaires, for both groups, related to changes perceived since the previous questionnaire was completed.

Associated with each version of the GUVQuest was a short clinician questionnaire, designed to provide clinical ratings and a range of other information for the purposes of identifying appropriate cases for inclusion in a longitudinal study. The clinician questionnaire also existed in four different versions for the uses described above. In each version the clinician was asked to award a pain score for the dog (on an 11-point NRS – with scoring from 0

to 10), and on follow-up questionnaires for dogs receiving treatment the clinician was also asked to provide an estimate of how the dog's chronic and painful condition had changed since the previous consultation. This rating was given on a 7-point Likert-type scale, with ratings chosen from the following descriptions: 'great deterioration', 'deterioration', 'slight deterioration', 'no change', 'slight improvement', 'improvement', and 'great improvement'.

4.2.1.2 Recruitment of subjects

The GUVQuest was completed by owners of dogs being treated for one of the following conditions, believed to be both chronic and painful: chronic degenerative joint disease, chronic otitis externa, anal furunculosis, painful tumours. The GUVQuest was also completed by the owners of a matched group of dogs (containing a similar range of breeds and ages, and including both male and female dogs) judged to be pain-free by UGSAH participating clinicians. For each owner questionnaire completed, the associated clinician questionnaire was also completed.

In accordance with the protocol shown in Appendix 11, eleven clinicians in three clinics of UGSAH – orthopaedic, oncology and soft tissue – were requested to recruit suitable cases (Hospital group), as was the veterinary practitioner operating an acupuncture clinic within a local veterinary practice (Practice group).

In addition, the questionnaires were completed by a group of owners of dogs that were not suffering chronic pain (Control group). Control group owners were recruited from among the staff and students of the University of Glasgow Faculty of Veterinary Medicine, who responded to an emailed request for volunteers.

4.2.1.3 Recruitment period

Recruitment of subjects was carried out over a period of approximately 18 months, between April 2001 and September 2002.

4.2.1.4 Recruitment criteria

Hospital group

Over the recruitment period, owners of all new cases for chronic orthopaedic conditions, painful tumours, chronic otitis externa and anal furunculosis attending UGSAH were asked by the relevant clinician to complete an initial owner questionnaire. The examining clinician also completed an initial clinician questionnaire. The information obtained from these questionnaires was used to provide data and to select for follow-up those owners/dogs who met the following criteria:

a dog was at least one year old;

b dog was diagnosed by examining clinician as being in some chronic pain caused by DJD, anal furunculosis, chronic otitis externa, or painful tumour;

c dog was likely to be seen again on at least two further occasions over the following 12 weeks;

d dog was likely to be seen on repeat visits by the clinician who completed the initial questionnaire;

e dog did not suffer any impairment such as poor eyesight, deafness, senility or physical handicap not associated with the condition of interest;

f dog had been owned by the person completing the initial questionnaire for at least one year **and** for longer than the owner believed the dog to have been in pain;

g all questions in the initial questionnaires (owner and clinician) had been answered.

h owner questionnaire was completed on the correct day.

Practice group

Owners of dogs attending an acupuncture clinic at a local veterinary practice were invited to participate in the study by completing an initial questionnaire, and thereafter were selected for participation in the longitudinal study using the same criteria as those selected for follow-up in the Hospital group.

Control group

Owners of pain-free dogs selected were all of those who volunteered by responding to an emailed request to all staff and students in the University of Glasgow Faculty of Veterinary Medicine, and whose dogs were subsequently found to be suffering no pain. An appointment was made for each volunteer to have his or her dog(s) examined by one of the UGSAH clinicians participating in the study. Owner and clinician completed an initial questionnaire from which owners of dogs assessed as having no pain (score of 0 on the 0-10 NRS pain scale included in the clinician's questionnaire) were recruited for the longitudinal study.

4.2.1.5 Administration, data handling and analysis

In all groups, every owner was given, with the initial questionnaire, a letter that explained the purpose of the study and the researcher's obligations under the data protection act. Each questionnaire was issued with a stamped addressed envelope. Owners were requested to complete the questionnaire at home on the day they received it, and to return it in the envelope provided. Owners and clinicians not returning questionnaires were contacted at least once by telephone, email or in person to request that the relevant questionnaire be returned.

For dogs in Hospital and Practice groups a follow-up questionnaire was issued at every consultation. For dogs in the Control group, a follow-up questionnaire was issued at a follow-up examination arranged to suit the examining clinician and owner.

The data capture from all returned questionnaires was carried out by author coding of the questionnaires and then manual data capture by a skilled data capture operator. All of the data thus captured was checked by the author against the coded questionnaires for accuracy, and corrections made to the data set where necessary.

Subsequent analysis of data was carried out using MINITAB for Windows® (Release 13).

4.2.2 Results

4.2.2.1 Recruitment

Of a total of 209 owners approached in the Hospital group and in the Practice group, 5 declined to take part in the study and a further 35 owners failed to return the initial questionnaire. Thus, an initial questionnaire was completed for a total of 169 dogs in both study groups. This represents a return rate of 82.8%.

Of the 155 Hospital group dogs who were recruited and whose owners returned their initial questionnaire, 52 were not enrolled for the longitudinal study, because they did not meet the requirements of the protocol. In order to recruit more cases, during the recruitment period the requirements **a**, **b** and **f** of the protocol were relaxed. Consequently, the age restriction was lowered to 3 months, resulting in the recruitment of 24 dogs aged <1 year into the Hospital group (for 11 of these dogs more than 1 questionnaire was completed: minimum age of dogs completing >1 questionnaire was 6 months) and 5 dogs aged <1 year into the Control group (for 4 of which dogs more than 1 questionnaire was completed: minimum age of dogs completing >1 questionnaire was 8 months). In addition, 18 dogs were recruited even though their owners answered 'no' to the question, 'do you think your dog is in any pain? (for 14 of which >1 questionnaire was completed – 2 of these dogs were also aged <1 year). The details of dogs in the Hospital group for which questionnaires were completed and data captured are given in Tables 4.1 and 4.2. Owners of 29 Hospital group dogs enrolled in the longitudinal study did not complete follow-up questionnaires for the following reasons: because the dog did not return to UGSAH (n=14), because the dog did return to UGSAH but in error the owner was not given a follow-up questionnaire (n=10), or because the owner did not return a follow-up questionnaire that was issued (n=5).

Table 4.1 Details of questionnaires completed and clinical conditions of Hospital group dogs enrolled on longitudinal study (n=103).

Key: DJD – degenerative joint disease; **PT** – painful tumour; **COE** – chronic otitis externa; **AF** – anal furunculosis; **ACPC** – another chronic and painful condition; **NCP** – no 'chronic pain' classification.

Of 103 dogs enrolled in longitudinal study, 76 dog owners completed ≥ 1 follow-up questionnaires	Of 103 dogs enrolled in longitudinal study, 29 dog owners completed 0 follow-up questionnaires

DJD	PT	COE	AF	ACPC	NCP	DJD	PT	COE	AF	ACPC	NCP
59	4	0	2	8	1	17	4	3	0	4	1

Table 4.2 Details of data capture and conditions of Hospital group dogs not enrolled on longitudinal study (n=52).

Key: DJD – degenerative joint disease; **PF** – painful tumour; **COE** – chronic otitis externa; **AF** – anal furunculosis; **ACPC** – another chronic and painful condition; **NCP** – no 'chronic pain' classification; **WQI** – wrong questionnaire issued to owner at initial visit; **MR** – too many missing responses.

Of 52 dogs not enrolled in longitudinal study, data was captured from 49 questionnaires	Of 52 dogs not enrolled in longitudinal study, data was not captured from 3 questionnaires

DJD	PT	COE	AF	ACPC	NCP	WQI	MR
18	8	1	1	12	9	2	1

4.2.2.2 Enrolment on longitudinal study for dogs suffering DJD and for Control group dogs

During the period of study, insufficient cases of painful tumours, chronic otitis externa and anal furunculosis were seen in order to permit the separate analysis of data obtained for each of these categories of chronic and painful condition. Analysis was therefore focussed on the group of dogs suffering from a range of conditions categorised by the examining clinician as DJD.

A total of 73 dogs with DJD provided at least one follow-up questionnaire: 59 dogs from the Hospital group (Hospital (DJD) group) and the 14 dogs in the Practice group, all of which were diagnosed with DJD.

The total number of questionnaires completed for the Hospital (DJD) group was 161 (range per owner 2-8, median questionnaires completed=2). The total number of questionnaires completed for the Practice group was 60 (range per owner 2-7, median questionnaires completed=4).

Twenty-three owners of 26 dogs participated in the longitudinal study as part of a Control group. Of the 26 participating dogs, an initial and one follow-up questionnaire were completed for 16 dogs (n owners=15), and only an initial questionnaire was completed for the remaining 10 dogs (n owners=8). Minimum period between examinations for Control group dogs was 56 days.

Demographic details for dogs with two or more questionnaires in Hospital (DJD), Practice and Control groups are shown in Appendix 12.

The data obtained from completed GUVQuests were analysed in studies designed to test and evaluate Hypotheses 1-6 that were presented on pages 110-111.

4.3 Assessment of validity of included measures of clinical status and clinical change

4.3.1 Testing Hypothesis 1

Hypothesis 1 was that clinician pain scores for dogs receiving treatment for a chronic and painful condition would tend to decrease with treatment for that condition.

4.3.1.1 Methods

An examination was made of the clinician pain scores awarded for each questionnaire (initial questionnaire was questionnaire 1, first follow-up questionnaire was questionnaire 2, and so on) completed for both Hospital (DJD) and Practice group dogs.

4.3.1.2 Results

Hospital (DJD) group

The results for this group are displayed in summary statistics (Table 4.3) and boxplot (Figure 4.1). Group pain scores showed a tendency to decrease from questionnaire 1 to questionnaire 2, but the median pain score was the same for questionnaire 2 and questionnaire 3, and the upper quartile extended higher for questionnaire 3 than for questionnaire 2.

The statistical significance of these changes was explored using the Wilcoxon Matched-pairs Signed-Ranks Test for Differences. The results of this test showed that the decrease in pain scores between questionnaires 1 and 2 (pain score 2 – pain score 1, where both pain scores were available) was statistically significant:

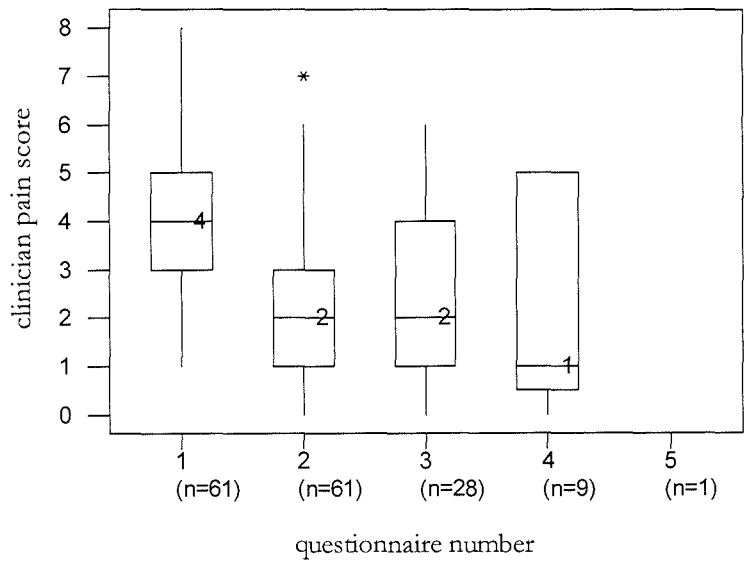
N pairs questionnaires	Median	Estimated Confidence	Achieved Confidence Interval
47	-1.50	95.0	(-2.00, -1.00)

Table 4.3 Hospital (DJD) group: summary statistics for clinician pain scores awarded using numerical rating scale (0-10).

Key: Quest No. – sequential questionnaire number; **N** – number of these questionnaires with clinician pain score; **N*** – number of these questionnaires with clinician pain score missing; **StDev** – standard deviation; **25th** – 25th percentile; **75th** – 75th percentile.

Quest No.	N	N*	Mean	Median	StDev	Min	Max	25th	75th
1	60	1	3.967	4.000	1.507	1.00	8.000	3.00	5.00
2	48	13	2.375	2.000	1.453	0.00	7.000	1.00	3.00
3	17	11	2.529	2.000	1.908	0.00	6.000	1.00	4.00
4	5	4	2.40	1.00	2.41	0.00	5.00	0.50	5.00
5	0	1	*	*	*	*	*	*	*

Figure 4.1 Hospital (DJD) group (n=59): boxplot showing clinician pain scores (on 0-10 NRS) for all questionnaires completed for this group.



Note: In Table 4.3 and Figure 4.1 the number of questionnaires 1 and 2 exceeds the number of dogs by 2, because two of the dogs had two separate courses of treatment, and completed a series of questionnaires for each course of treatment.

However, the changes in pain scores between questionnaires 2 and 3 (pain score 3 – pain score 2, where both pain scores were available) were not statistically significant, since the confidence interval includes 0:

N pairs questionnaires	Median	Estimated Confidence	Achieved Confidence Interval
16	-0.50	94.8	(-2.00, 0.50)

Practice group

With the Practice group, a similar pattern of decreasing pain scores over time was seen, as shown in the summary statistics (Table 4.4) and boxplot (Figure 4.2) for clinician pain scores for this group. Group pain scores showed a tendency to decrease from questionnaire 1 to questionnaire 3, and, overall, from questionnaire 1 to questionnaire 7, but with a small increase from questionnaire 3 to questionnaire 4. The statistical significance of the changes recorded over the first 4 questionnaires was explored using the Wilcoxon Matched-pairs Signed-Ranks Test for Differences.

The results of applying this test to differences between questionnaires 1 and 2 (pain score 2 – pain score 1) where both pain scores were available, were:

N pairs questionnaires	Median	Estimated Confidence	Achieved Confidence Interval
12	-0.50	95.0	(-1.50, 0.00)

The results of applying the test to differences between questionnaires 2 and 3 (pain score 3 – pain score 2) where both pain scores were available, were:

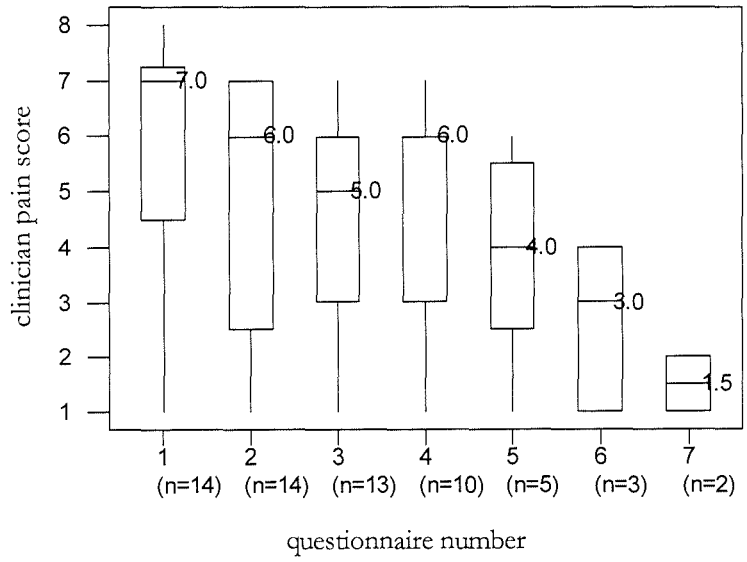
N pairs questionnaires	Median	Estimated Confidence	Achieved Confidence Interval
10	-1.00	94.5	(-1.50, -0.50)

Table 4.4 Practice group: summary statistics for clinician pain scores awarded using numerical rating scale (0-10).

Key: Quest No. – sequential questionnaire number; **N** – number of these questionnaires with clinician pain score; **N*** – number of those questionnaires with clinician pain score missing; **StDev** – standard deviation; **25th** – 25th percentile; **75th** – 75th percentile.

Quest No.	N	N*	Mean	Median	StDev	Min	Max	25th	75th
1	14	0	6.000	7.000	2.184	1	8	4.50	7.25
2	13	1	5.000	6.000	2.273	1	7	2.50	7.00
3	13	0	4.538	5.000	2.025	1	7	3.00	6.00
4	9	0	4.889	6.000	2.261	1	7	3.00	6.00
5	5	0	4.000	4.000	1.871	1	6	2.50	5.50
6	3	0	2.667	3.000	1.528	1	4	1.00	4.00
7	2	0	1.500	1.500	0.707	1	2	*	*

Figure 4.2 Practice group (n=14): boxplot showing clinician pain scores (on 0-10 NRS) for all questionnaires completed for this group.



Finally, the results of applying the test to the differences between questionnaires 3 and 4 (pain score 4 – pain score 3) where both pain scores were available, were:

N pairs questionnaires	Median	Estimated Confidence	Achieved Confidence Interval
9	0.00	95.6	(-0.50, 0.50)

The results of the the Wilcoxon Matched-pairs Signed-Ranks Test for Differences between pairs of questionnaires available for dogs in the Practice group therefore showed that the changes in pain scores awarded between questionnaires 1 and 2 and between questionnaires 3 and 4 were not statistically significant, while the reduction in pain scores between questionnaires 2 and 3 was significant.

Consequently, the results for both Hospital (DJD) and Practice groups confirm the hypothesis that clinician pain scores for dogs receiving treatment for a chronic and painful condition tended to decrease with treatment for that condition.

4.3.2 Evaluating Hypothesis 2

Hypothesis 2 was that there would be a relationship between a clinician's pain scores and that clinician's assessment of clinical change at each follow-up clinical examination.

4.3.2.1 Methods

For every pair of consecutive questionnaires completed for dogs in the Hospital (DJD) and Practice groups, with the relevant questions completed, an examination was made of the concordance between the pain scores awarded on the consecutive occasions and the assessment of clinical change awarded on the second occasion. The data were obtained from the clinician's estimate of clinical change (Clinician Follow-up Questionnaire, question 9), and from the clinician's pain scores (Clinician Questionnaires, question 1)

awarded on consecutive occasions. Each of the clinicians' estimates of change and each pair of pain scores was placed in one of 3 categories according to whether these suggested that the dog had improved, deteriorated, or there had been no change. Concordances between these data were then examined and further categorised as 'agreement', as 'disagreement' or as 'rating of change in opposite direction to direction of change in pain scores', as follows:

'Agreement' – when the direction of change in the pain scores agreed with the assessment of clinical change or if no change in the pain score was matched by a clinical assessment of 'no change';

'Disagreement' – when no change in pain scores was associated with a clinical assessment of some improvement or deterioration, or if some change in pain scores was associated with a clinical assessment of 'no change';

'Rating of change in opposite direction to the direction of change in pain scores' – when the direction of change in pain scores was the opposite of that indicated by the assessment of clinical change.

4.3.2.2 Results

Hospital (DJD) group

Table 4.5 shows the percentage of occasions (on which the clinician made consecutive examinations and completed relevant questions: n=59) on which each level of concordance between a clinician's pain scores and assessment of clinical change was noted for this group (n clinicians=6).

This data revealed agreement on approximately 75% of occasions between the clinician's pain scores and his or her estimate of change in the dog's chronic and painful condition since its previous examination. On approximately 25% of occasions the direction of change of pain scores and the clinicians' assessment of change did not agree, and on 6 of these occasions (*c.*10%) an estimate of change in one direction was recorded where the clinician's pain scores indicated change in the opposite direction.

Table 4.5 Table to show % occasions on which a clinician's rating of change agreed or did not agree with that clinician's pain scores on the relevant occasions. For Hospital (DJD) group: n clinicians=6; n estimates of change/pain scores available=59. For Practice group: n clinicians=1; n estimates of change/pain scores available=36.

Group of dogs to which ratings were given	**% estimates of change that disagreed with clinician's pain scores**	**% estimates of change that agreed with clinician's pain scores**	**% estimates of change that were opposite to direction of change in pain scores**
Hospital (DJD) group	15.25	74.58	10.17
Practice group	19.44	77.78	2.78

Practice group

Results from the Practice group (n clinicians=1) are also shown in Table 4.5.

For 36 pairs of questionnaires with the relevant questions completed (n occasions=36), there was agreement on approximately 78% of occasions between the clinician's estimate of the way in which the dog's chronic and painful condition had changed since last seen and the direction of change in clinician pain scores awarded on the relevant occasions. On approximately 22% of occasions the direction of change of pain scores and the clinician's assessment of change did not agree. In most cases this was where no change in one was recorded as a change in the other, but in one of these cases an estimate of change in one direction was recorded where the clinician's pain scores indicated change in the opposite direction.

Considered together, these data showed agreement between clinicians' pain scores and their estimates of change on more than 7 out of 10 occasions where both ratings were available, but some disagreement on approximately 25% of occasions.

4.3.3 Evaluating Hypothesis 3

Hypothesis 3 was that 'yes' answers to the owners' question 'do you think your dog is in any pain?' would be associated with higher clinician pain scores than would 'no' answers to that question.

4.3.3.1 Methods

Because this evaluation did not require longitudinal data, in order to obtain a larger sample, questionnaires from all dogs in the Hospital group that were suffering from DJD were examined, including those for dogs not enrolled on the longitudinal study. The pain scores data for dogs in this group and in the Practice group were examined separately according to whether the owner had answered 'yes' or 'no' to the relevant question.

4.3.3.2 Results

Hospital group

The number of questionnaires included in this analysis was 293 (n dogs=94). The median pain score given to dogs whose owners answered 'no' was 2 (inter-quartile range of 1-3) and that given to dogs whose owners answered 'yes' was 4 (inter-quartile range of 3-5).

Practice group

The number of questionnaires included in this analysis was 58 (n dogs=14). Clinician pain scores for dogs whose owners answered 'no' had a median pain score of 4 (inter-quartile range of 1-6) while those answering 'yes' had a median pain score of 6 (inter-quartile range of 1-8).

The data for both groups showed that higher pain scores were associated with dogs whose owners reported that their dogs were in pain ('yes' responses) compared to dogs whose owners reported that they were not ('no' responses).

4.3.4 Evaluating Hypothesis 4

Hypothesis 4 was that ratings of change would more frequently be awarded to transition questions by owners of dogs with chronic pain than by owners of Control group dogs.

4.3.4.1 Materials and methods

There were 8 transition questions in the owners' questionnaire, which required owners to rate global change in the behavioural domains of activity, pain, sociability, aggression, anxiety, enthusiasm, happiness and mobility. In the initial questionnaire owners were asked to rate change since their dogs became unwell. In follow-up questionnaires owners were asked to provide a rating of change since the previous questionnaire was completed. The ratings from which owners were able to choose to respond to these transition questions were as follows: 'greatly decreased' (GD); 'decreased' (D); 'slightly decreased' (SD); 'no change' (NC); 'slightly increased' (SI); 'increased' (I); 'greatly increased' (GI).

Bar charts illustrating the % owners awarding each response for each transition question were prepared in order to examine the relative frequency of these global reports for each domain. Such bar charts were prepared for 4 groups of questionnaires:

Group 1 This group consisted of all initial questionnaires completed for dogs in the Hospital (DJD) group, to reveal changes observed with the onset of a chronic and painful condition;

Group 2 This group consisted of all questionnaires (initial and follow-up) for dogs in the Hospital (DJD) group, to reveal the areas in which most change was observed over time with onset of and treatment for a chronic and painful condition;

Group 3 This group consisted of all follow-up questionnaires for dogs in the Control group, to reveal the extent of any changes observed in healthy dogs over time.

4.3.4.2 Results

Group 1: Initial questionnaires for Hospital (DJD) group

The ratings given by owners to the transition questions in initial questionnaires, which were completed before treatment commenced, recorded owners' perceptions of change in their dogs since the dogs had become unwell.

The bar charts that follow (Figure 4.3a-h) show the relative frequency of ratings awarded for each transition question in the initial questionnaires completed for the Hospital (DJD) group (n dogs=59). Most owners rated activity, enthusiasm, happiness and mobility decreased, and pain increased, since their dogs became unwell, and little change in sociability, aggression or anxiety. Those who did report change in anxiety mostly reported it increased.

Figure 4.3a-h Hospital (DJD) group: bar charts showing relative frequency of responses to each transition question in the initial questionnaire (GD, greatly decreased; D, decreased; SD, slightly decreased; NC, no change; SI, slightly increased; I, increased; GI, greatly increased).

Figure 4.3a Hospital (DJD) group, all dogs with at least one follow-up questionnaire: bar chart showing % of owners awarding each rating for change in activity since dog became unwell (n=59)

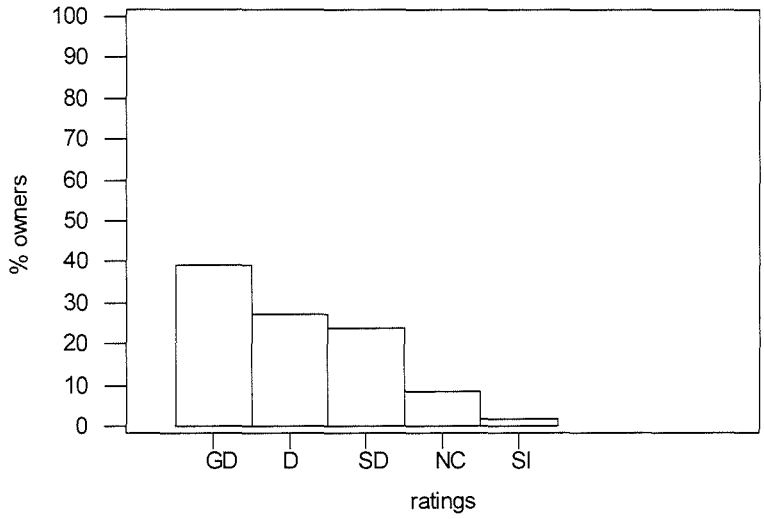


Figure 4.3b Hospital (DJD) group, all dogs with at least one follow-up questionnaire: bar chart showing % of owners awarding each rating for change in pain since dog became unwell (n owners=59)

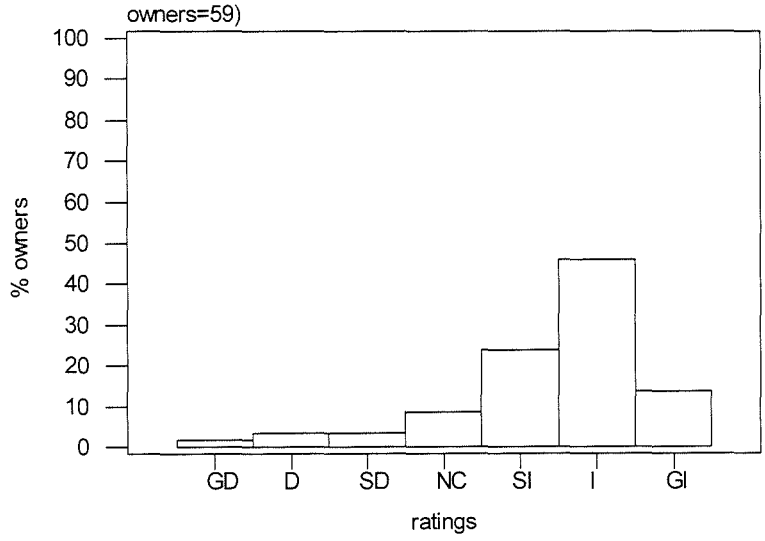


Figure 4.3c Hospital (DJD) group, all dogs with at least one follow-up questionnaire: bar chart showing % of owners awarding each rating for change in sociability since dog became unwell (n owners=59)

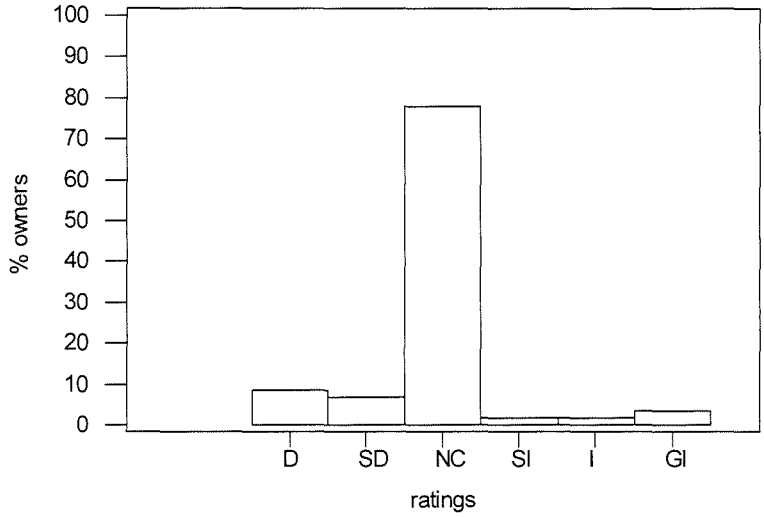


Figure 4.3d Hospital (DJD) group, all dogs with at least one follow-up questionnaire: bar chart showing % of owners awarding each rating for change in aggression since dog became unwell (n owners=59)

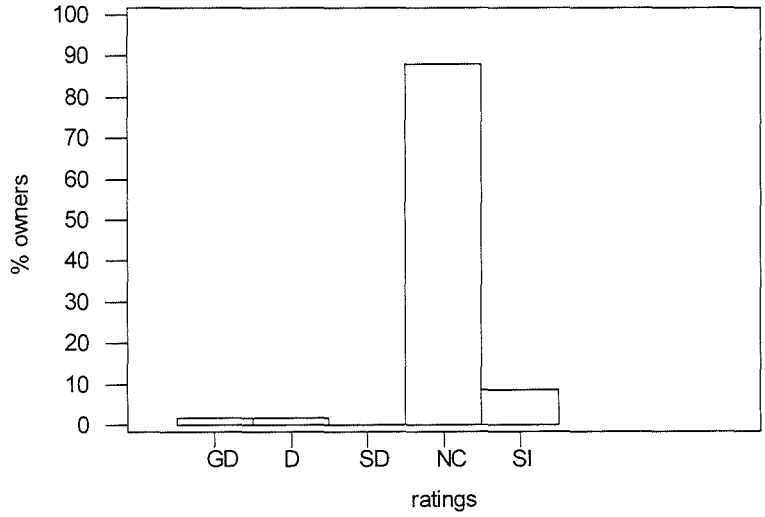


Figure 4.3e Hospital (DJD) group, all dogs with at least one follow-up questionnaire: bar chart showing % of owners awarding each rating for change in anxiety since dog became unwell (n owners=59)

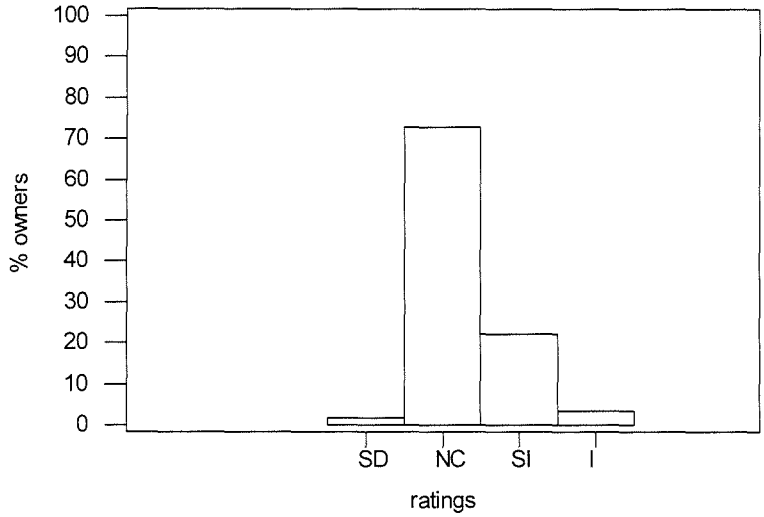


Figure 4.3f Hospital (DJD) group, all dogs with at least one follow-up questionnaire: bar chart showing % of owners awarding each rating for change in enthusiasm since dog became unwell (n owners=59)

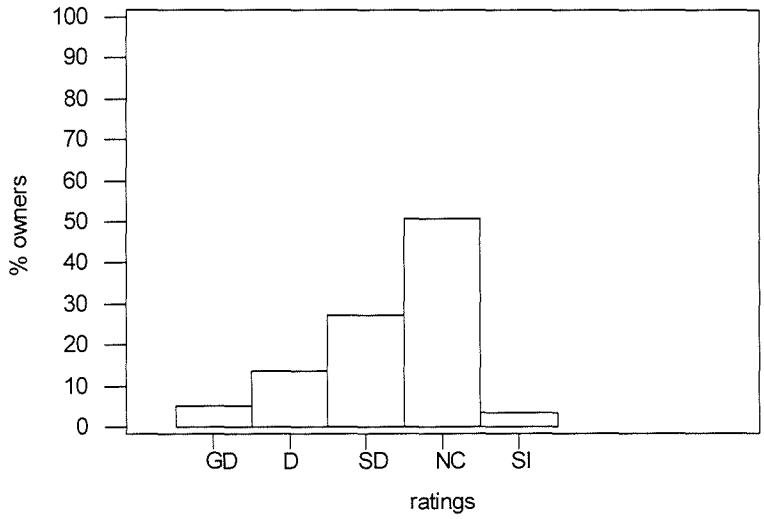


Figure 4.3g Hospital (DJD) group, all dogs with at least one follow-up questionnaire: bar chart showing % of owners awarding each rating for change in happiness since dog became unwell (n owners=59)

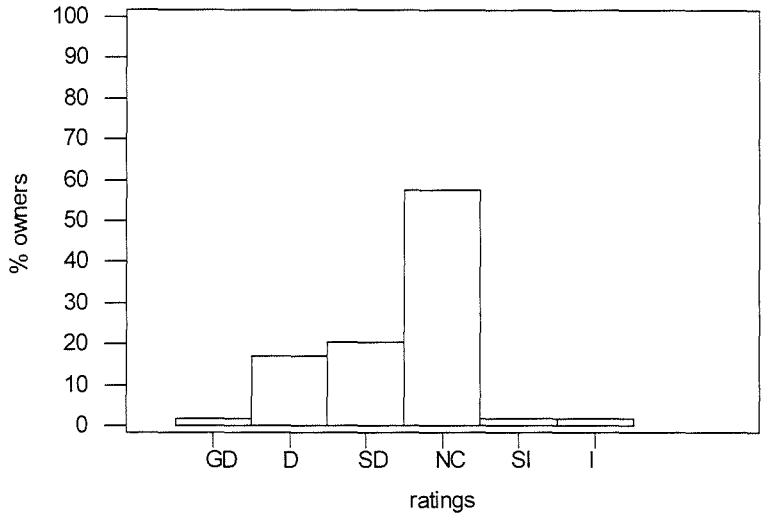
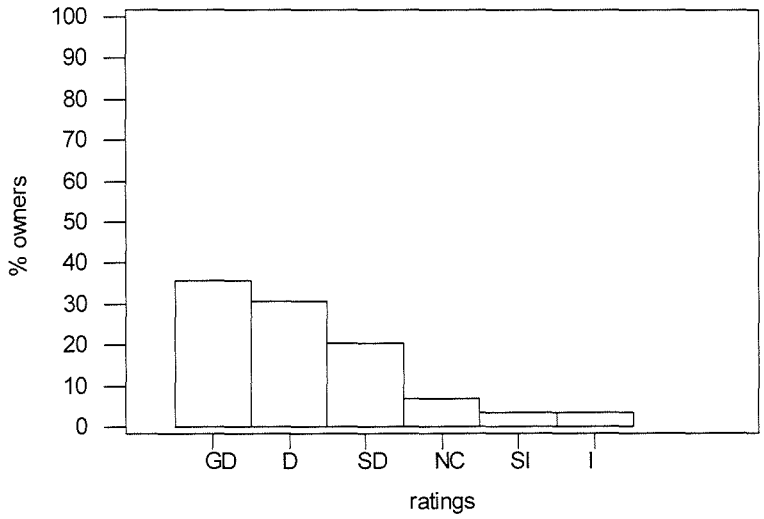


Figure 4.3h Hospital (DJD) group, all dogs with at least one follow-up questionnaire: bar chart showing % of owners awarding each rating for change in mobility since dog became unwell (n owners=59)



Group 2: All questionnaires for Hospital (DJD) group

The responses to transition questions in all questionnaires (initial and follow-up questionnaires) given by owners of dogs in the Hospital (DJD) group are shown in Figure 4.4a-h.

With the onset of a dog's chronic and painful condition, and during treatment for that condition, owners in the Hospital (DJD) group reported very little change in aggression, a little more change in sociability and anxiety, still more change in enthusiasm and happiness, and most change in activity, pain and mobility.

Group 3: Follow-up questionnaires for Control group

For owners of dogs in the Control group, transition questions were only present on follow-up questionnaires. By contrast with the transition question ratings for Hospital (DJD) and Practice groups, the owners of dogs in the Control group (n dogs=16; n follow-up questionnaires=16) reported no change for most transition questions. One owner reported slightly decreased sociability at follow-up, 1 reported slightly increased aggression, and 1 reported both slightly reduced sociability and slightly increased aggression at follow-up. All other reports were of no change.

These results demonstrate that, in response to transition questions, owners of dogs with chronic pain provided ratings indicating some change in behaviour for most transition items whereas owners of Control group dogs reported no change for most transition items.

4.3.5 Evaluating Hypothesis 5

Hypothesis 5 was that, for dogs in the Hospital (DJD) and Practice groups, associations would be seen between ratings for owner transition questions, since the behavioural domains that these addressed were hypothesised to be affected by chronic pain (expecting positive associations between ratings of activity, sociability, enthusiasm, happiness and mobility, and negative associations between these and ratings of pain, aggression and anxiety).

Figure 4.4a-h Hospital (DJD) group: bar charts showing relative frequency of responses to each transition question in all questionnaires (GD, greatly decreased; D, decreased; SD, slightly decreased; NC, no change; SI, slightly increased; I, increased; GI, greatly increased).

Figure 4.4a Hospital (DJD) group: bar chart to show % of occasions each rating chosen by owners for change in activity, in all questionnaires generated by this group (n=161)

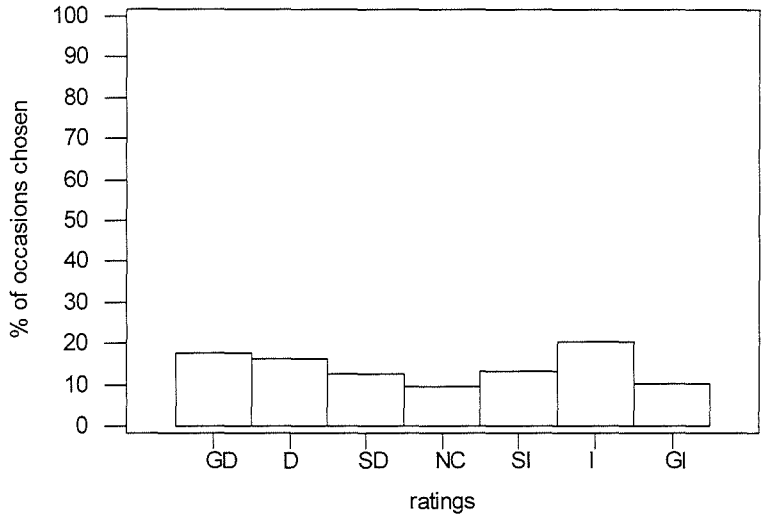


Figure 4.4b Hospital (DJD) group: bar chart to show % of occasions each rating chosen by owners for change in pain, in all questionnaires generated by this group (n=161)

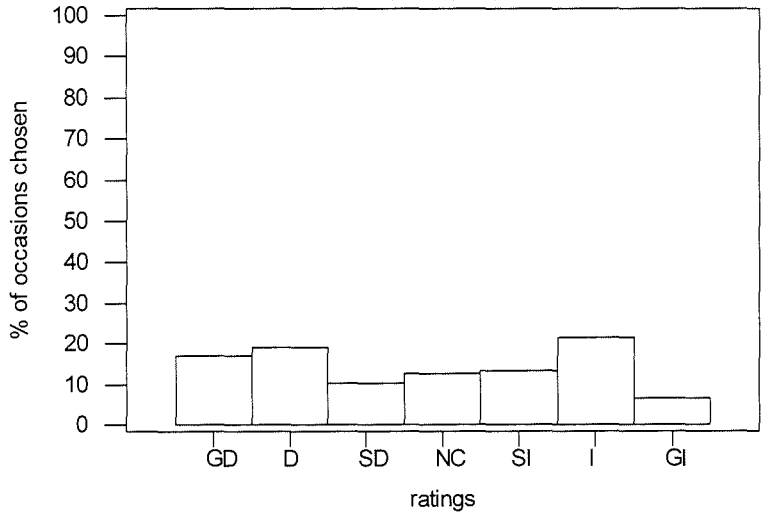


Figure 4.4c Hospital (DJD) group: bar chart to show % of occasions each rating chosen by owners for change in sociability, in all questionnaires generated by this group (n=161)

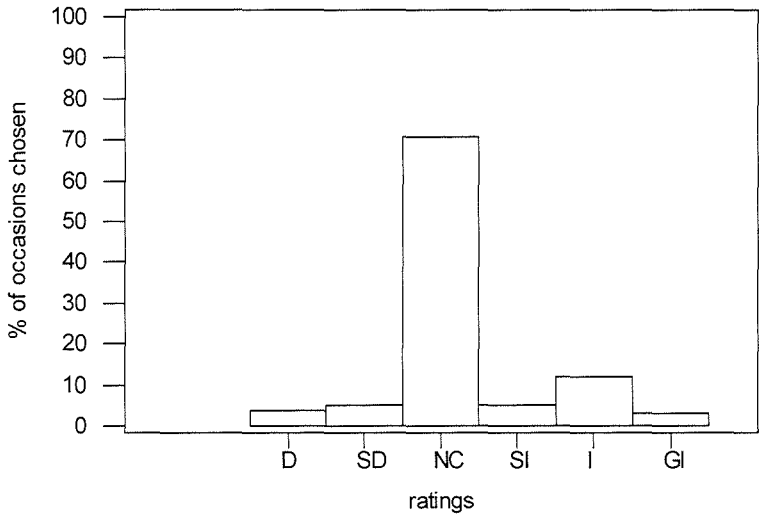


Figure 4.4d Hospital (DJD) group: bar chart to show % of occasions each rating chosen by owners for change in aggression, in all questionnaires generated by this group (n=161)

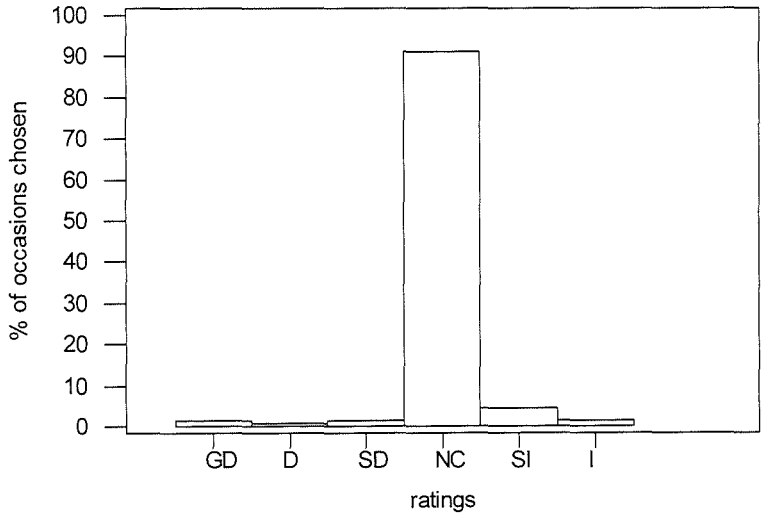


Figure 4.4e Hospital (DJD) group: bar chart to show % of occasions each rating chosen by owners for change in anxiety, in all questionnaires generated by this group (n=161)

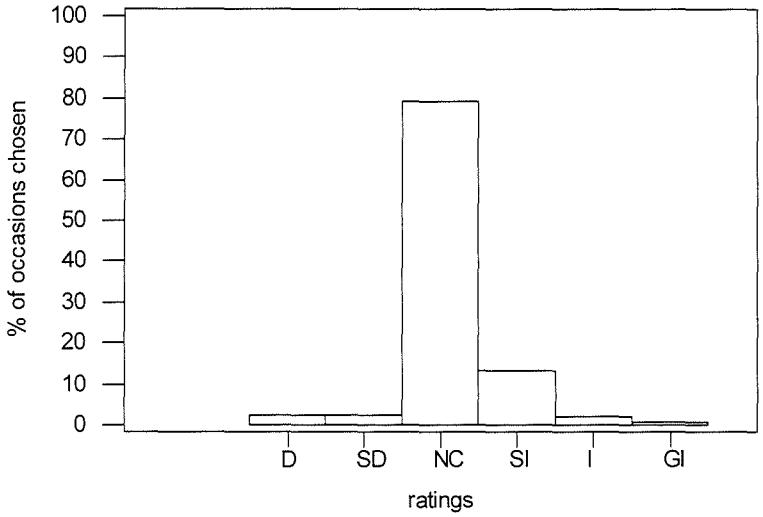


Figure 4.4f Hospital (DJD) group: bar chart to show % of occasions each rating chosen by owners for change in enthusiasm, in all questionnaires generated by this group (n=161)

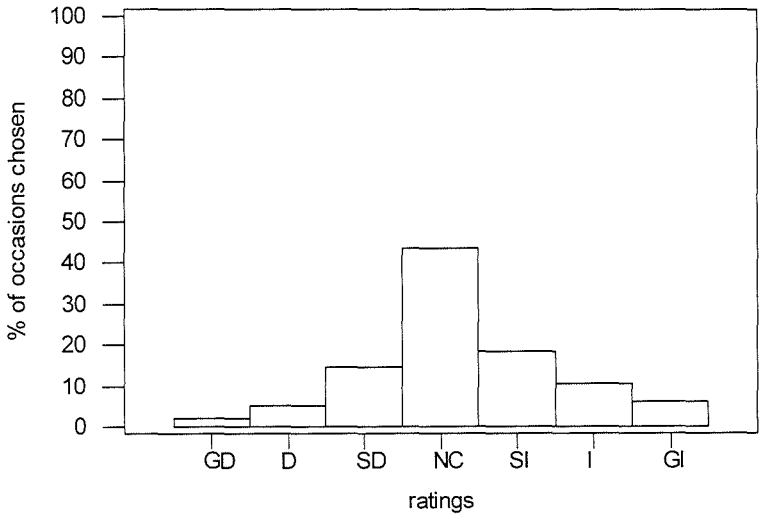


Figure 4.4g Hospital (DJD) group: bar chart to show % of occasions each rating chosen by owners for change in happiness, in all questionnaires generated by this group (n=161)

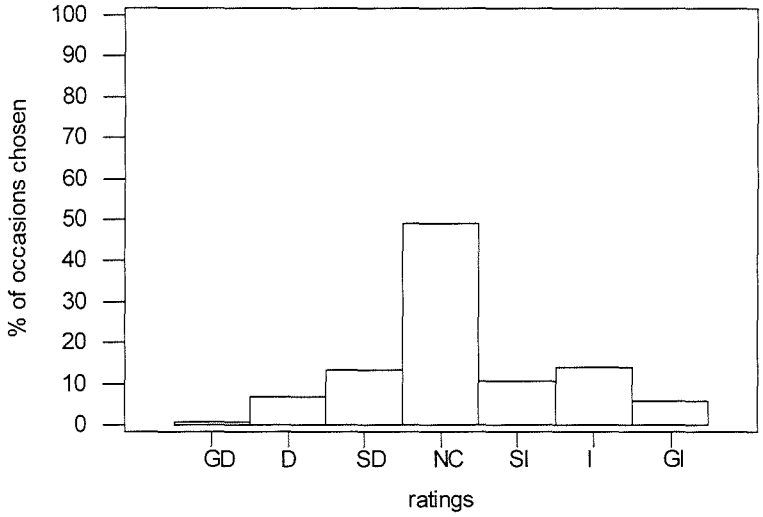
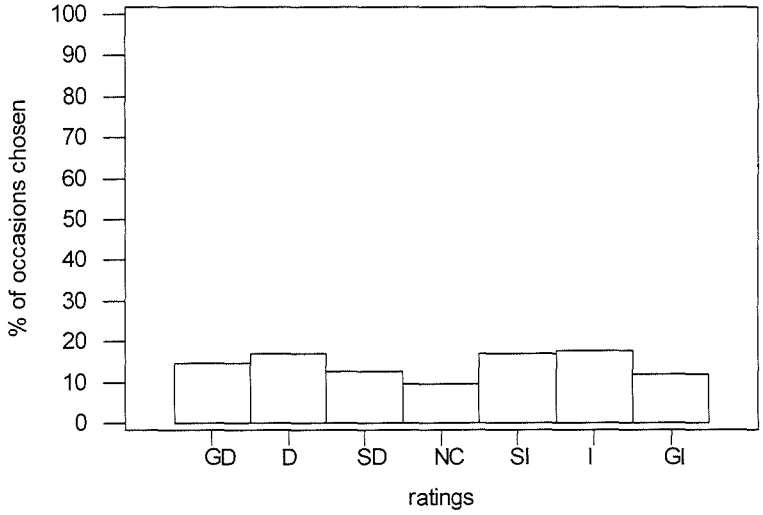


Figure 4.4h Hospital (DJD) group: bar chart to show % of occasions each rating chosen by owners for change in mobility, in all questionnaires generated by this group (n=161)



4.3.5.1 Methods

The ratings for each transition question were cross-tabulated with every other transition question, separately, for all questionnaires completed for the Hospital (DJD) group and for the Practice group.

4.3.5.2 Results

Hospital (DJD) group

The cross-tabulations of transition question ratings for all questionnaires for Hospital (DJD) group dogs with multiple questionnaires are shown in Appendix 13, and one example is provided in Table 4.6.

For this group, reported changes in pain generally showed a negative association with reported changes in activity, though this was not as clear as might have been expected. Reported changes in pain were also largely negatively associated with changes in sociability. However, many reports (approximately 70%) were of no change in sociability, even when pain was reported as greatly increased or decreased. There was a clear negative association between changes in pain and changes in enthusiasm, happiness and mobility, and some indication of a positive association between pain and aggression.

Generally positive associations were found between ratings of change in activity, sociability, enthusiasm, happiness, and mobility.

Most reports were of no change in anxiety (78% of reports) or in aggression (89% of reports). Where changes were reported (22% reports rated some change in anxiety, and 11% rated some change in aggression), those for aggression showed no clear associations with changes in activity, enthusiasm or happiness, although there did appear to be a negative association with changes in mobility. The few changes in anxiety that were reported appeared mostly to be negatively associated with changes in activity, enthusiasm, happiness and mobility, and positively associated with changes in pain.

Table 4.6 Cross-tabulation of transition question ratings for owners' ratings of change in ACTIVITY (rows) and owners' ratings of change in PAIN (columns), with data taken from all questionnaires for dogs in Hospital (DJD) group.

(GD): greatly decreased; (D): decreased; (SD): slightly decreased; (NC): no change; (SI): slightly increased; (I): increased; (GI): greatly increased; *: no rating given

	*	(GD)	(D)	(SD)	(NC)	(SI)	(I)	(GI)	All
*	3	0	0	0	0	0	0	0	**3**
(GD)	0	2	4	1	0	2	13	6	**28**
(D)	0	2	2	1	1	5	12	3	**26**
(SD)	0	1	2	2	0	7	7	1	**20**
(NC)	0	0	3	0	8	3	1	0	**15**
(SI)	0	2	5	10	3	1	0	0	**21**
(I)	0	10	11	2	6	2	1	0	**32**
(GI)	0	10	3	0	2	1	0	0	**16**
All	**3**	**27**	**30**	**16**	**20**	**21**	**34**	**10**	**161**

Since there were very few ratings of change in aggression or anxiety, it was difficult to see any associations between the ratings of these two transition questions. For the same reason, the few reports of change in anxiety showed no clear association with changes in sociability. However, the few changes that were reported did appear to indicate a negative association between sociability and aggression.

Practice group

Cross-tabulations of the ratings of each transition question with every other, for all questionnaires for the Practice group (see Appendix 14) revealed some of the expected associations between reported changes in each. Reported changes in pain generally showed a negative association with reported changes in activity, enthusiasm, happiness and mobility. Similarly, negative associations were apparent between changes in pain and in sociability, although more than half of ratings of sociability were of 'no change', even where pain was rated as decreased or increased (though not greatly so).

Generally positive associations were shown between ratings of change in activity, sociability, enthusiasm, happiness, and mobility.

There were very few ratings of change in aggression or anxiety. Consequently, it was difficult to see any associations between the ratings of these two transition questions and even with the other transition questions. However, the few reports of change in anxiety did appear to be negatively associated with reports of changed sociability, enthusiasm, happiness and mobility. Unexpectedly, the few ratings of change in anxiety or aggression were generally positively associated with changes in activity.

4.3.6 Evaluating Hypothesis 6

Hypothesis 6 was that owner responses to transition questions, at least those to which owners most frequently responded with a rating of some degree of change, would demonstrate expected associations with clinician assessments of clinical change. For example, it was expected that clinicians' reports of

improvement in a dog's chronic and painful condition would be associated with owner reports of increased activity, sociability (where change was reported), enthusiasm, happiness and mobility, and decreased pain and (where change was reported) anxiety. The converse was expected when a clinician reported deterioration in a dog's chronic and painful condition.

Although not clear from the results of evaluating Hypotheses 4 and 5, it was also expected, from the results of interviews with owners and of descriptor-generating exercises carried out as part of this study, that any reported change in aggression would be negatively associated with clinicians' reports of improvement and positively associated with their reports of deterioration.

4.3.6.1 Methods

For dogs in the Hospital (DJD) and Practice groups, clinicians' assessments of change in a dog's chronic and painful condition were cross-tabulated with owners' ratings of change in activity, pain, sociability, aggression, anxiety, enthusiasm, happiness and mobility.

It should be noted that while the letters used to code clinician and owner transition questions were the same, the meaning of the coding for owner ratings of transition questions was different from that for the clinician assessments of change.

An interpretation of the coding for each type of question should be made as follows:

	Clinician assessment of change	*Owner transition questions*
GD	Great Deterioration	Greatly Decreased
D	Deterioration	Decreased
SD	Slight Deterioration	Slightly Decreased
NC	No Change	No Change
SI	Slight Improvement	Slightly Increased
I	Improvement	Increased
GI	Great Improvement	Greatly Increased

4.3.6.2 Results

Hospital (DJD) group

There is no estimate of clinical change in the clinician's initial questionnaire. Thus, in order to reduce the number of missing clinician ratings, only the follow-up questionnaires were included in cross-tabulations of clinicians' assessments of change and owners' transition question ratings. There were 98 follow-up questionnaires completed for dogs in the Hospital (DJD) group. The cross-tabulations of ratings from these, and interpretations of those cross-tabulations, are given in Appendix 15. An example of the cross-tabulations is given in Table 4.7, and was interpreted as follows. Of the 58 clinician assessments available, 49 rated the dog's condition as improved to some degree. With a clinician's assessment of improvement, 32 owner ratings were of increased activity. Similarly, 2 assessments of clinical deterioration were matched by owner ratings of decreased activity, but for 3 clinical assessments of deterioration the owner ratings were of increased activity. On 2 occasions there was a clinical assessment of no change while the owner rated activity as increased, and on 5 occasions the owner rated activity as unchanged while clinical assessment was of improvement.

The Hospital (DJD) group cross-tabulations revealed that where changes were reported, these were generally associated in the expected manner with clinician reports of improvement or deterioration in the dog's clinical condition, with improvement associated with owner reports of increase in activity, enthusiasm, happiness and mobility, and to a lesser extent sociability, and with owner reports of decrease in pain.

Practice group

Cross-tabulations of clinicians' assessments of change in condition with owners' transition question ratings, for all questionnaires completed by the Practice group (n=60), are given in Appendix 16. There were no clinician reports of any deterioration over time in the condition of any dogs within this group.

Table 4.7 Cross-tabulation of clinicians' assessments of change (rows) with owners' ratings of change in ACTIVITY (columns) with data taken from follow-up questionnaires completed for dogs in Hospital (DJD) group.

Key:

	Rows	*Columns*
(GD)	Great deterioration	Greatly decreased
(D)	Deterioration	Decreased
(SD)	Slight deterioration	Slightly decreased
(NC)	No Change	No change
(SI)	Slight Improvement	Slightly increased
(I)	Improvement	Increased
(GI)	Great Improvement	Greatly increased
*	missing rating	missing rating

	*	(GD)	(D)	(SD)	(NC)	(SI)	(I)	(GI)	All
*	0	0	5	1	5	7	17	5	**40**
(D)	0	0	1	0	0	0	0	0	**1**
(SD)	2	0	1	0	0	2	1	0	**6**
(NC)	0	0	0	0	0	1	0	1	**2**
(SI)	0	1	1	1	2	3	6	3	**17**
(I)	1	2	0	3	3	4	6	4	**23**
(GI)	0	1	1	1	0	2	2	2	**9**
All	**3**	**4**	**9**	**6**	**10**	**19**	**32**	**15**	**98**

These results showed that, in general, when the clinician assessed that there had been clinical improvement, owners reported increased activity, sociability, enthusiasm, happiness and mobility, and decreased pain, as was predicted. Few owners reported any change in aggression or in anxiety, so no association with the clinicians' assessment of change was apparent for those transition questions (for a range of clinicians' estimates of change, 'no change' in aggression and in anxiety was reported by most owners).

4.4 Discussion

Where a population is being sampled it is important to maximise compliance, since missing questionnaires can result in biased results (Vaillancourt *et al.*, 1991). Among all dog owners approached in UGSAH and in the veterinary general practice to participate in this study, a number (n=40) were either initially unwilling to participate (n=5) or expressed themselves willing but failed to return the initial questionnaire issued to them (n=35). This represents a response rate of >80%. No information was gathered about the non-respondents' characteristics that would permit the identification of similarities or differences between respondents and non-respondents. This was unfortunate, since such comparison may have permitted greater confidence that the sample of owners who did respond was representative of the population from which they were sampled. The response rate of >80% for the first questionnaire is comparable with that reported in a recent study to assess the feasibility, reliability and validity of the parent form of the Child Health Questionnaire (CHQ-PF50) (Raat *et al.*, 2002), in which, of 467 questionnaires distributed, 364 were returned. This 78% response rate was considered by the authors to contribute to an 'excellent' feasibility performance.

In the development of proxy instruments for the assessment of human chronic pain and HRQL, it is usual to compare the results obtained using a new proxy instrument with those obtained by self-report (Sprangers and

Aaronson, 1992) or with an existing 'gold standard' for the construct being measured. Unfortunately, confirmatory self-report is not possible when developing a proxy instrument for animal pain or HRQL, nor is there currently an existing 'gold standard' against which to measure such a new instrument. Consequently, it was necessary in this study to include, alongside the instrument's core items, additional measures with which to compare the scores obtained from ratings of those core items, by means of which to categorise known groups (e.g. chronic pain and no pain) and to assess clinical change: the clinicians' pain scores and estimate of change, and the owners' transition question ratings. These measures themselves, however, required validation. In this chapter a series of hypotheses were evaluated in order to assess the validity of these additional measures for the purposes of validating, subsequently, the instrument's core items.

Clinician pain scores for both Hospital (DJD) and Practice groups largely followed the expected pattern over time, which supported Hypothesis 1 – that clinician pain scores would decrease with treatment for a chronic and painful condition – and provided some evidence for the validity of the clinicians' pain scores as an indicator of clinical change.

The level of agreement between the clinicians' assessments of change and the pain scores awarded, for both Hospital (DJD) and Practice groups, suggested that there was some validity in both of those measures (supporting Hypothesis 2), although there was a degree of disagreement between them in approximately 25% of the questionnaires in these samples. Even if clinicians were basing pain scores on an attribute such as lameness, which is possible, particularly for orthopaedic cases, agreement would still be expected between changes in pain scores and the associated estimates of change, so the lack of concordance between these measures was surprising, but may simply be a consequence of the difficulty of recalling individual cases over long periods of time. The time between assessments varied widely among the pairs of questionnaires, from 5 days to 135 days in the Practice group and from 9 days

to 298 days in the Hospital (DJD) group. (For Control group dogs, minimum time between questionnaires was 56 days and maximum was 81 days.) For the Hospital (DJD) and Practice group dogs, the time between questionnaires was determined by timing of hospital or practice appointments, according to clinical criteria. For Control group dogs, the time between questionnaires was determined according to the convenience of clinician and owner, with a minimum time between questionnaires (6 weeks) determined by the author.

Although instructions to both owner and clinician requested that their respective questionnaires be completed on the same day as the physical examination, in practice some respondents (both owners and clinicians) completed their questionnaires at a later time, and so their responses may have been influenced by factors relating to accuracy of recall of the dog on the day of the examination.

An alternative explanation for the discrepancy between clinicians' pain scores and estimates of change may be that real but subtle clinical changes are not sufficiently large to be reflected in a changed pain score. For example, one of the dogs in the Practice group was given a score of 1 at every visit (low level of pain from first visit). While improvements in the dog's condition may have been apparent to the clinician, there may still have been some residual pain, so that it was not possible to award a pain score of 0 even though some improvement was noted. It may be that the method of clinician pain scoring adopted here (and commonly used) is not sufficiently sensitive to capture subtle changes in the dog's painful condition.

To eliminate the problem of inter-observer variability (where the reports of different proxies may not be equivalent), in longitudinal studies of proxy measurement of paediatric HRQL it is recommended that the same proxy rater be used throughout (Matza *et al.*, 2004). For the same reason, the GUVQuest required to be completed by the same owner on each occasion throughout the longitudinal study.

In both Practice and Hospital (DJD) groups, owners' answers of 'yes' to the question, 'do you think your dog is in any pain?' were associated with higher clinician pain scores than those that were awarded to dogs for which the owners answered 'no' to that question, which supports a degree of validity for both measures. However, the results from both groups suggested that owners tended to underestimate or under-report pain, or that clinicians tended to over-report pain: many owners answered 'no' to the question 'do you think your dog is in any pain?' while the clinician awarded the dog a pain score of at least 1 (and a score as high as 3 for the Hospital group and as high as 6 for the Practice group). Any under-reporting by owners may be accounted for by some owners being poor at recognising when their dogs were in pain, or because the word 'pain' suggested a more strongly aversive experience than some owners believed their dogs to be suffering (and for such owners the term 'discomfort' may have seemed more appropriate), or because some owners were reluctant to contemplate that their dogs may have been in any pain and may have been in such pain for some time. It should be noted that owners answered this question (on whether or not they believed pain to be present) after their consultations, during which they are likely to have discussed with the clinician the possibility of the presence of pain. In these circumstances, any under-reporting of pain by the owner might be considered to be very strongly motivated.

An alternative explanation would be that owners did not recognise relevant behaviour changes as being associated with pain, because of their insidious onset. This was remarked upon by the veterinary specialists and veterinary general practitioners interviewed in the early stages of this study, and has been noted by others (Flecknell, 1985; Mathews, 2000; Robertson, 2003).

Owners of dogs with chronic degenerative joint disease reported global changes in behavioural domains, over time, that were not reported by owners of dogs in the Control group that did not have such conditions, as was hypothesised (Hypothesis 4). This provided evidence that the transition

questions were revealing behaviour changes associated with chronic pain and so had validity for eliciting global behavioural assessments that were associated with change in chronic pain status.

Hypothesis 5 was supported by the data since, taken together, the transition question ratings of owners of dogs in Hospital (DJD) and Practice groups largely revealed expected associations between transition questions, although the predicted changes in aggression and anxiety and their relationship with changes in pain were not so apparent. The fairly clear negative associations between changes in pain and changes in activity, sociability, enthusiasm, happiness, and mobility, may be accounted for in part by the likelihood that the owner is making deductions about the change in pain from the other perceived changes. The generally positive associations between ratings of change in activity, sociability, enthusiasm, happiness and mobility, may be accounted for in part by the relative ease with which sociability, enthusiasm and happiness may be demonstrated when a dog is more mobile and active.

The associations between the clinicians' estimates of change and owners' transition questions supported Hypothesis 6, providing evidence for the validity of the clinicians' estimates of change and for the owners' ratings of transition questions regarding activity, pain, sociability, enthusiasm, happiness and mobility. However, there were some unexpected results. For the Hospital (DJD) group, in a few cases where the clinician assessed that the dog's condition had deteriorated, the owner ratings were of increased activity. One explanation for this might be an increase in restlessness with increased pain (as reported in semi-structured interviews described in Chapter 2), giving rise to a report of increased activity. In both Hospital (DJD) and Practice groups, reports of changes in aggression and anxiety were fewer than had been expected. There are several possible explanations for this: it may reflect a lesser influence of chronic pain upon these types of behaviour, a lesser sensitivity on the part of the owner to changes in these kinds of behaviour, or an owner's reluctance to report changes in anxiety or aggression because these

are seen as undesirable traits. This is a problem of respondent bias, which was introduced in Chapter 3 (page 107). It would be interesting to compare the ratings of aggression from owners of companion dogs with owners of, for example, police dogs for which a degree of aggression may be considered to be desirable.

In many cases where the clinician assessed that there had been some degree of clinical change, owners reported 'no change' in sociability. This may reflect some reluctance on the part of owners to recognize or to report any reduction in sociability (or increased sociability, which would imply a reduced sociability in the past) because sociability is perceived as a desirable trait in companion dogs. Alternatively, it may be that sociability, which has been selected for in most breeds, is a relatively inelastic trait, so that changes will be seen only when the pain is severe or very prolonged. In most cases, any report of increased sociability was associated with a clinician report of improvement. In a few instances, owners of dogs in the Hospital (DJD) group reported increased sociability when clinician reports were of clinical deterioration, which may be accounted for by an increase in help-seeking behaviour with increased pain.

Unexpectedly, in the Practice group the ratings reporting change in anxiety or aggression were generally positively associated with changes in activity, perhaps reflecting an increase in restlessness, or a return to 'normal' levels of aggression as activity increased. The few ratings of change in aggression were also positively associated with changes in mobility, suggesting that it may be a return to normal function that permits 'normal' aggressive behaviour. However, where only small numbers of owners reported any change, any perceived associations must be viewed very cautiously.

The analysis described in this chapter provided some evidence for the validity of clinicians' pain ratings and assessments of clinical change. It also provided some evidence for the validity of owners' transition question ratings

particularly regarding pain, mobility, activity, enthusiasm and happiness and, to a slightly lesser extent, sociability.

The validation of the instrument itself required the exploration by statistical analysis of the ratings given by owners to the core descriptor items. If owners were using the 109 descriptors and their associated scales to provide information about a range of behavioural domains that were relevant to the measurement of chronic pain, then statistical associations between the item ratings would be expected to reveal an underlying structure that was interpretable as the construct upon which the instrument was developed. In this case, the instrument was intended to measure chronic pain, and it was expected to do so by measurement of its impact upon a range of domains of canine HRQL. If such an underlying structure was revealed by analysis, then those domains of HRQL would provide HRQL domain scores which could be compared with the various measures of clinical status and clinical change that were validated to some extent in this chapter. These are aspects of construct validation (see Chapter 1, page 15), the process of which will be more fully described, and the results of which will be reported, in the following chapter.

Chapter 5

CONSTRUCT VALIDATION BY FACTOR ANALYSIS OF INSTRUMENT RESPONSES AND BY USE OF INSTRUMENT SCORES

5.1 Introduction

Validation may be thought of as a process of hypothesis testing, to answer the question 'does the hypothesis of this validation study make sense in light of what the scale is designed to measure?'(Streiner and Norman, 1995). One important type of validity is construct validity. There are a variety of ways in which evidence for construct validity may be sought, and this chapter describes the ways in which such evidence was sought for the GUVQuest, some of which required the development of a practicable scoring method for the instrument.

5.1.1 Construct validity

In psychiatry, the trait that is being measured is generally not itself visible, but is inferred from a variety of observations. The trait exists only as a hypothetical construct, which must be tested (Streiner, 1993). Evidence for the construct validity of an instrument is provided when the responses obtained with that instrument fit the hypothetical construct upon which the instrument was developed.

5.1.1.1 Using factor analysis to obtain evidence for construct validity

Factorial validity is one type of construct validity, which requires the statistical analysis of relationships between responses to the items of an instrument. Groupings of items revealed by such analysis, that are also related on clinical or other grounds, are termed 'factors'. If an interpretable factor structure underlies the responses to instrument items, and if this underlying factor structure fits the construct upon which the instrument was developed, then

some evidence has been provided for the validity of the instrument and also for that hypothetical construct (Johnston, 1998).

Factor analysis (FA) is a multivariate statistical technique that is designed to reveal an otherwise 'hidden' structure underlying the responses to an instrument's various items, by identifying a number of underlying relationships (factors) between the variables (instrument items), and identifying which items belong to each factor.

Common variance is the variance that a variable shares with other variables; unique variance is the variance that is unique to a particular variable. The principal components method of FA analyses the total variance (both common and unique) of the variables and seeks a linear combination of the variables that extracts maximum variance from them. This variance is then removed and a second linear combination of variables is sought that accounts for as much as possible of the remaining variance, and so on. A useful factor model is one that captures a reasonable amount of the total variance, with higher figures representing better models.

Because FA is capable of providing, for a given data set, any number of factor models, it is for the instrument developer to decide upon the most satisfactory factor model and the number of factors it contains, which is a vital step in instrument development (Coste *et al.*, 2005). There are various established methods by which this decision may be reached. In FA, the eigenvalue is an important figure attached to each of the potential factors, which indicates how much of the variance is accounted for by a given factor. One method of choosing the most suitable factor model is to use a 'scree test' in which a graph of eigenvalues, forming a scree plot, is used to decide how many factors are required in order best to represent the data. The number of factors required is that number where the slope of the scree graph changes markedly, beyond which each factor accounts for much less of the variance than the factors appearing before this 'elbow'. Another method of selecting

the number of factors to be extracted uses the Kaiser criterion, which is that all factors with an eigenvalue greater than 0.1 should be included in the model. It has been suggested that the scree test may result in too few factors being selected (Statsoft, 1984-2003) and that the use of the Kaiser criterion tends to result in too many factors being extracted (Coste *et al.*, 2005). Coste and colleagues (2005) recommend that a careful and diversified approach be taken to determining the number of factors to retain, and favour the use of confirmatory factor analysis (in the sense in which this term is used by statisticians, to mean a repetition of the analysis on a new sample; cf. page 154) to confirm the results of an initial FA.

Importantly, a good factor model is one in which the statistically derived factors are interpretable (ACITS, The University of Texas at Austin, 1995-97). With fast modern methods of statistical analysis, FA can be performed with various values for the number of factors to be extracted, and a model then selected that is the most sensible on clinical or other grounds (Darlington, n.d.). While a larger number of factors will account for more of the variance, factors defined by only one ('singlet') or two ('doublet') observed variables are not considered desirable (ACITS, The University of Texas at Austin, 1995-97).

When carrying out the FA, an aid to interpretation of the factor structure is to use factor rotation. The purpose of this is to maximise the loading of a variable on one factor, while minimising its loading on all other factors. Where the factors are expected to be uncorrelated, an orthagonal rotation should be used: varimax is a commonly used orthogonal rotation (Nunally, 1994).

All interpretations of factors should be regarded as tentative, subject to confirmation by further research (Friendly, 1995). Such confirmation may be provided in a single study by randomly splitting a sample into two, and using one half for exploratory analysis and the other for confirmatory. There are

two accepted meanings of the term 'exploratory factor analysis'. One, used by statisticians, describes the initial factor analysis on a component of a data set, with 'confirmatory factor analysis' describing factor analysis on the remaining component of the data set, with the intention of confirming the results of the exploratory analysis. However, instrument developers commonly use the term 'confirmatory factor analysis' to describe a factor analysis designed to confirm a hypothesised factor model. For example, it was reported that the core scales of the PedsQL™4.0 displayed a factor structure that was largely consistent with the *a priori* conceptually derived scales (Varni *et al.*, 2001). Conversely, a FA of responses to the SF-36 identified only 6 meaningful factors whereas the instrument was hypothesised to measure 8 dimensions (Wolinsky *et al.*, 1998).

For each factor model, each item will be associated to some extent with the underlying factors. This association is expressed in an item's factor loading, with higher loadings representing closer associations. Factor loadings above 0.3 or above 0.4 are generally considered to be moderate or high (Burgess, 2001; North Carolina State University, n.d.), and those above 0.6 may be considered to be high or very high, depending on the type of scale associated with an item (with higher loadings required for Likert scales compared with dichotomous scales) (Burgess, 2001). Loadings may be positive or negative but the signs of the loadings are relatively, not absolutely, important. The largest loadings provide an indication of the identity of the factor, and zero or low loadings can confirm that identity. Ideally, each item would load significantly onto only one factor but in practice it is not uncommon for an item to have multiple significant loadings. However, if there are any items that fail to load significantly onto any factor, then the analyst may consider attempting to derive a new factor solution after excluding them (ACITS, The University of Texas at Austin, 1995-97).

In addition to an item's factor loading, important information is also provided by an item's communality. This is calculated for each item by summing its squared loadings on the factors. Communality figures range from 0 (none of

the variability is accounted for by the underlying factor model) to 1.0 (all of the variability is accounted for by the underlying factor model). The communality of an item for a chosen factor model is a measure of how much of that item's variability is accounted for by the underlying factor model, and therefore provides an indication of the model's adequacy (with regard to explaining the variance of each item) (Ackerley, n.d.).

The number of subjects required for a satisfactory FA is generally large, with some experts advocating a sample size of between 5 and 10 cases for each variable, and a minimum of 150 cases being suggested (Burgess, 2001; Canadian Forest Service, 2002). However, there is no absolute criterion, as the adequacy of the sample size depends to some extent upon the properties of the data, the number of factors extracted and the size of the correlations (Burgess, 2001). The clearer the true factor structure, the smaller the sample size needed to reveal it.

Factor analysis has been used in the development of a number of human pain and, particularly, HRQL instruments. The factor models revealed in this way have accounted for a range of variances. A 5-factor structure for the Pain Attitudes Questionnaire (revised) accounted for around 56% of the variance (Yong *et al.*, 2003); a 4-factor structure for an infant quality of life questionnaire accounted for some 45% of the variance (Manificat *et al.* 1999); 5-factor structures for self-report and proxy-report responses to items in the Pediatric Quality of Life Inventory™ Version 4.0 accounted for 52% and 62% of the variance, respectively (Varni *et al.* 2001); a 5-factor model for the Asthma Quality of Life questionnaire accounted for around 53% of the variance (Juniper *et al.* 1997); a 5-factor structure for the 20-item Neck Pain and Disability scale accounted for 76% of the variance (Wheeler *et al.*, 1999); and FA of the 16-item Chronic Heart Failure Questionnaire revealed a 3-factor structure that accounted for 68% of the variance (Wolinsky *et al.*, 1998).

Factor analysis has also been applied to instruments that measure the temperament of dogs, using questionnaires designed to obtain relevant information from owners. Factor analysis of a questionnaire designed to evaluate in this way the behaviour and temperament of guide dogs, extracted 8 factors that together accounted for 63% of the common variance in item scores (Serpell and Hsu, 2001), and the factor analysis of a more recent questionnaire designed to measure the behaviour and temperament of pet dogs revealed an 11-factor structure that together accounted for 57% of the common variance (Hsu and Serpell, 2003).

A range of factor solutions, ranging from 2-factors to 9-factors, emerged from a series of studies to investigate the factor structure of the Rotterdam Symptom Checklist (Fayers and Hand, 2002), and a variable number of factors have also been revealed by factor-analytic studies of the MPQ (Prieto and Geisinger, 1983). It has been suggested that such results are caused by the inclusion in an instrument of items addressing indicator variables (variables that reflect but do not influence the subject of measurement) and causal variables (variables that may have an influence on the subject of measurement). It has been recommended that psychometric scales should contain only items related to indicator variables whereas the items of clinimetric scales may also address one or more causal variables, and that scale developers must be alert to the implications of combining these (Fayers and Hand, 2002).

5.1.1.2 Additional evidence for construct validity

Apart from factorial validity, other approaches to construct validation depend upon the extent to which the performance of an instrument reflects the hypothetical construct upon which it was developed. This approach to validation is one in which predictions are made about how scores obtained with the new instrument will differ between groups, after treatment, over time, or relate to other measures of change, and these predictions are then tested. In order to test the performance of an instrument in this way, it is

necessary to develop a scoring method for the instrument in which the responses to individual items may be translated into a score or a set of scores.

The scores obtained with a new instrument can be used in various ways to obtain evidence for the construct validity of that instrument. For example, an instrument should be able to distinguish correctly between groups that would be expected to have quite different scores on that instrument. This is called extreme groups (Johnston, 1998) or known groups (Fayers and Hand, 2002) validity. This approach was used to validate the five-core-cues PICIC (Pain Indicator for Communicatively Impaired Children) instrument, which correctly classified 87.4% of pain/no pain episodes (Stallard *et al.*, 2002). Discrimination between pain and no-pain situations in populations of infants was also used as a measure of the construct validity of the EDIN (Debillon *et al.*, 2001). The PedsQL™4.0 generic core scales demonstrated significant differences between a healthy population and paediatric rheumatic diseases groups (Varni *et al.*, 2002a; Varni *et al.*, 2002b), and the validity of the CHQ and the HUI2 (Health Utilities Index 2) were compared by examining their ability to discriminate between a group of children without any chronic condition and a group with at least two (parent-reported) chronic conditions, and also by comparing a group of children with no medical consumption and one with at least three visits to the doctor in the previous year (Raat *et al.*, 2002). Finally, the validation of the Pediatric Cancer Quality of Life Inventory–32 (PCQL–32) used a known groups discrimination approach, and 3 of its 5 subscales were found to discriminate between groups on treatment versus those off treatment (Varni *et al.*, 1998).

Such tests provide evidence of construct validity and demonstrate an instrument's validity for discriminative purposes. However, clinicians will more usually require an instrument to be valid for evaluative purposes. That is, it has to be able to detect change over time within a patient (Fayers and Hand, 2002). An instrument that is responsive to clinical change would be expected to generate scores that would reflect clinical improvement or

deterioration over time. For example, the PedsQL™4.0 examined changes in scores over time, with clinical intervention, as an indicator of instrument responsiveness (Varni *et al.*, 2002a; Varni *et al.*, 2002b). Similarly, evidence for the construct validity of the Royal Marsden Hospital Paediatric Quality of Life Questionnaire was obtained by comparing ratings of HRQL at baseline and 6-8 weeks after treatment commenced (Watson *et al.*, 1999), and developers sought to validate the Non-Communicating Children's Pain Checklist by comparing ratings given before and after surgery (Breau *et al.*, 2002a).

In the case of the GUVQuest, the hypothetical construct upon which the instrument was developed would lead to the prediction that scores for a group of dogs receiving treatment for a chronic and painful condition would be different from the scores for healthy dogs. Furthermore, it would be expected that, with treatment, the scores of dogs in a chronic pain group would change over time to more closely resemble the scores obtained for healthy dogs (which would be expected to be relatively stable over time).

The purpose of the GUVQuest was to evaluate change in individuals, with individuals acting as their own controls. Different breeds and ages of dogs, and even different individuals, may be expected to have baseline scores that differ from those of other breeds, ages or individuals. In these circumstances it was particularly important to examine the way in which changes in scores for individual dogs reflected clinical change in that individual. Furthermore, it was possible that variables such as age, sex or prior experiences of owners might influence their ratings, since similar influences have been suggested for human proxy instruments (Levi and Drotar, 1999; Eiser and Morse, 2001). For that reason, it was considered to be important that the questionnaires from which the scores were obtained were completed by the same rater on each occasion, as recommended for longitudinal studies involving proxy raters of paediatric HRQL (Matza *et al.*, 2004).

The following hypotheses relating to the construct validity of the GUVQuest were made and subsequently tested and evaluated.

Hypothesis 1

It was hypothesised that an underlying factor model would be revealed by FA of the instrument item responses and that such factors would be interpretable as a range of domains of HRQL likely to be affected when a dog is suffering chronic pain. Given the similarity between the impacts of chronic pain on dogs and people reported in this study, it was further hypothesised that the HRQL domains identified by this analysis would be similar to the range of domains included in human HRQL instruments.

Hypothesis 2

It was hypothesised that scores obtained with the instrument would be able to discriminate between healthy dogs and dogs with a clinical diagnosis of a chronic and painful condition, or between dogs with a clinician-awarded pain score of 0 and dogs with a pain score >0.

Hypothesis 3

It was hypothesised that group level comparisons of scores obtained with the instrument would reveal clear differences between those obtained for Control group dogs and for those dogs that were diagnosed with a chronic and painful condition (Clinical group). Differences were predicted between the range of scores obtained with the instrument for dogs in the Control group compared with dogs in the Clinical group, with the range of scores obtained for Control group dogs expected to be narrower than that for dogs in the Clinical group. Differences were also predicted in the stability of such scores for these two groups over time: that group level scores for Control group dogs would remain fairly stable over time, whereas group level scores for Clinical group dogs would change over time (assuming a reduction of chronic pain with treatment) from scores that were different from those obtained for Control

group dogs, towards scores that were more similar to those obtained for Control group dogs.

Hypothesis 4

Finally, it was hypothesised that the scores for individual cases would reflect clinical change (whether improvement or deterioration) or clinical stability over time.

5.1.2 Development of a scoring method

Responses to items provide, in themselves, a set of scores generated by an instrument. However, where the number of items included in an instrument is large, such scores are impracticable. Where a single attribute is being measured it may be appropriate to translate the responses for all items into a single score. However, where measurement is of a multi-dimensional construct, such as that of pain or HRQL, the amalgamation of scores for different domains or dimensions may result in the loss of valuable information, and might even produce an overall score that is meaningless. Most HRQL instruments are designed to generate a set of scores for a subject (Matza *et al.*, 2004), each score relating to one domain or dimension of HRQL, such a collection of domain scores providing a profile of scores for the individual. In some cases, subscale scores may be summed to form scores for broader dimensions of the construct being measured. For example, the Child Health Questionnaire (Landgraf *et al.*, 1998) may be scored in two ways: summing the ratings for items associated with each 'concept' provides a 14-concept profile; the concept scores can then be aggregated to derive a score for physical and one for psychosocial health (Medical Outcomes Trust, 2001).

Because the results of FA can be strongly influenced by the presence of error in the original data, it has been recommended that rather than using statistically derived factor scores as domain scores, summated scores should be constructed directly from item responses, which provide the added benefit of preserving the variation in the data (Web Centre for Social Research

Methods, 2004). Since different factors may include different numbers of items, it is common to use an average instead of a total score for each factor contributing to the measurement (Fayers and Hand, 2002). In cases where factors include both positive and negative items, it is usual to manipulate the item responses so that these do not cancel one another out. For example, in the development and validation of a questionnaire for measuring behaviour and temperament traits in pet dogs, Hsu and Serpell (2003) generated a score for each factor by calculating the mean of the scores for all items for that factor, after reversing scores for items with negative loadings for that factor.

Other scoring methods include that adopted for the Neck Pain and Disability Scale, which was to sum the responses to items 'loading heavily' (loadings >0.5) onto each of the multi-item factors identified (Wheeler *et al.*, 1999). Alternatively, the PedsQL™4.0 scores are computed as the sum of items divided by the number of items answered, which accounts for missing data (no score computed if >50% of item scores are missing) (Varni *et al.*, 2002a).

5.2 Methods and results

5.2.1 Testing Hypothesis 1

Factor analysis was used to identify an underlying factor structure and this was compared with the hypothetical construct upon which instrument was developed.

5.2.1.1 Methods

In order to have sufficient cases to investigate a factor structure statistically, a relatively large data set is required. The data set obtained in the course of this study was insufficiently large to allow it to be split in two in order to carry out simultaneously an exploratory and a confirmatory (in statistical parlance) FA. The single FA carried out may be considered, in the parlance of instrument development, to be a confirmatory analysis, in the sense that it was designed to confirm the hypothesised construct upon which the instrument was developed.

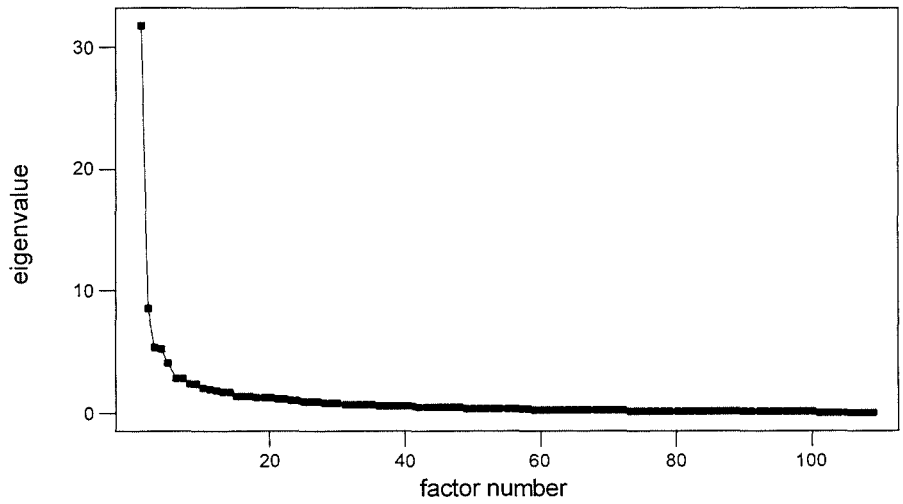
Because the instrument was intended primarily to be an evaluative tool, rather than a discriminative one, it was decided to perform the FA only on data obtained from dogs that went on to complete more than one questionnaire. The dataset used for this analysis was therefore that obtained for all dogs in the Hospital (DJD) and Practice groups: those that were suffering DJD and for which there had been completed more than one questionnaire (n questionnaires=221; n dogs=73). Multivariate Factor analysis was carried out on this dataset, using MINITAB for Windows® (Release 13). A principal components method of FA was used, and a varimax rotation performed. Input variables were all descriptor ratings. Loadings were sorted, and loadings of less than 0.3 were zeroed.

A scree test and the Kaiser criterion were both used to obtain an estimate of the approximate number of factors that would be likely to provide a suitable factor model. Guided by these tests, the interpretability of a range of factor models was then examined. Factors were interpreted on the basis of the descriptors loading onto and not loading onto a particular factor, and how those descriptors were related. In order to obtain an indication of the number of factors that would account for an acceptable amount of the variability in the dataset, FA was repeated for a range of values for the number of factors to extract, and the percentage of the variance accounted for by each model was examined. The factor model chosen was the one that accounted for an acceptable amount of the variability in the data, was most readily interpretable, and did not include factors containing only one or two items.

5.2.1.2 Results

The scree plot is shown in Figure 5.1. Although this showed no marked change of slope – no 'elbow' – its shape suggested that much of the variance in the dataset was accounted for by 7 factors, and that more than 20 factors accounted for little additional variance.

Figure 5.1 Scree plot of eigenvalue against factor number calculated from dataset of descriptor ratings for all questionnaires completed for dogs in Hospital (DJD) and Practice groups.



The Kaiser criterion (that all factors with an eigenvalue greater than 0.1 should be included in the model) suggested that a model containing around 15 factors would be appropriate.

A careful examination of the items loading onto each factor was made, and the consequent interpretability considered, for a range of factor models: from a 7-factor model to a 15-factor model. The amount of variability accounted for by a range of factor models (Table 5.1) was also considered.

While all of the factor models examined were interpretable to some degree, the most appropriate appeared to be the 12-factor model, which accounted for over 65% of the variability in the data. The sorted rotated factor loadings for the 12-factor model obtained from the Hospital (DJD) group and Practice group datasets is shown in Appendix 17. The communalities for this model are given in numerical order in Appendix 18.

Table 5.2 shows the 12 factors that were interpreted from this model. Each factor was named, as far as possible, after two items loading heavily and only onto that factor (onto no other factor). The items loading onto each factor (with minimum loading of 0.3) are grouped according to whether they had positive or negative loadings, although the signs of these loadings are only relatively and not absolutely important. It can be seen from the table that 8 factors contained both positively- and negatively-loading items, and 4 factors contained only positively-loading or only negatively-loading items. The loadings (positive or negative) of most descriptors relative to other items loading onto the same factors accorded with their prior classification as being either 'positive descriptors' or 'negative descriptors'. Positive descriptors were terms that owners had used in previous phases of the study to describe a healthy dog, and negative descriptors were terms that owners had used to describe a dog in chronic pain. Eight of the 12 factors contained both positive descriptors and negative descriptors. Factors 3, 6 and 9 contained only negative descriptors, and Factor 12 contained only positive descriptors.

Table 5.1 Table to show amount of variability in the data accounted for by various factor models, revealed by a range of factor analyses performed on data obtained from all dogs in Hospital (DJD) and Practice groups (n questionnaires=221; n questionnaires with missing values=56; n questionnaires used=165).

Number of factors in model	**% variance accounted for by that factor model**
1	29.1
2	37.0
3	41.9
4	46.6
5	50.4
6	53.0
7	55.6
8	57.8
9	59.9
10	61.9
11	63.6
12	65.2
13	66.8
14	68.3
15	69.6

Table 5.2 Table of named factors identified in an interpretable 12-factor model, showing items (descriptors) loading positively or negatively onto each factor. Order in which items are listed indicates weight of loading, with earlier listing indicating higher loading.

Factor number and name	**Items (descriptors) with –ve loadings loading onto that factor**	**Items (descriptors) with +ve loadings loading onto that factor**
Factor 1 Eager–keen	Eager, keen, inquisitive, energetic, outgoing, curious, lively, bouncy, bold, excitable, bright, boisterous, playful, nosy, alert, active, interested, fun-loving, sociable, stretching, confident, comfortable, athletic, fit, relaxed, contented, happy, easy-going, independent	Quiet, slowed, tired, lethargic, lacklustre, sluggish, weary, subdued
Factor 2 Stiff–sore	Stiff, sore, limping, pained, uncomfortable, awkward, slowed, resigned, tired, apprehensive, miserable, weary, pathetic/pitiful, unhappy, agitated, restless, unsettled, distressed, sorrowful, sad	Energetic, lively, bouncy, boisterous, playful, active, comfortable, athletic, fit, relaxed, contented, happy
Factor 3 Listless–reluctant	Slowed, resigned, miserable, lethargic, listless, lacklustre, reluctant, sluggish, apathetic, weary, sleepy, depressed, dull, subdued, withdrawn, pathetic/pitiful, unhappy, unsociable, detached, sorrowful, sad, uninterested	

Table 5.2 continued from page 166.

Factor number and name	**Items (descriptors) with –ve loadings loading onto that factor**	**Items (descriptors) with +ve loadings loading onto that factor**
Factor 4 Panicky–nervous	Apprehensive, panicky, nervous, uneasy, frightened, upset, strained, anxious, cautious, agitated, panting, restless, distressed	Confident, easy-going, calm, laid-back
Factor 5 Aggressive–unresponsive	Sociable, good-natured, even-tempered, friendly, affectionate, easy-going, placid	Withdrawn, unhappy, agitated, aggressive, irritable, grumpy, unsociable, compulsive, unresponsive, restless, territorial-protective
Factor 6 Whining–crying	Compulsive, whining, crying, moaning, groaning, panting, disturbed, restless, unsettled, picky (food), off his/her food, complaining	
Factor 7 Enthusiastic about food–interested in food	Picky (food), off his/her food, sorrowful	Enthusiastic about food, interested in food, greedy, tireless
Factor 8 At ease–consistent	Anxious, agitated Inconsistent	Quiet, contented, good-natured, friendly, easy-going, consistent, calm, placid, laid-back, obedient, at ease
Factor 9 Confused–complaining		Pained, miserable, depressed, dull, irritable, grumpy, moaning, groaning, confused, complaining, distressed
Factor 10 Attention-seeking–comfort-seeking	Affectionate, attention-seeking, comfort-seeking, clingy, thirsty	Independent, detached.
Factor 11 Sorrowful–sad	Inquisitive, happy, territorial-protective	Unhappy, sorrowful, sad, uninterested
Factor 12 Stoical–accepting		Thirsty, stoical, accepting

However, 'territorial-protective' had been classified *a priori* as a positive descriptor but had loaded with negative descriptors onto Factor 5 (Aggressive–unresponsive). There was similar disagreement between the *a priori* classifications and the factor loadings for the descriptors 'affectionate', 'thirsty' and 'detached' within Factor 10 (Attention-seeking–comfort-seeking).

Each of the factors in the 12-factor model contained items that were interpreted as bearing some relation to one another, and the name given to each factor reflected the nature of the domain addressed by that factor. Thus, Factor 1 (Eager–keen) contained items describing aspects of vitality or lack of vitality, and Factor 2 (Stiff–sore) contained items that appeared to be related to physical limitations or activity.

Only two items appeared to be completely unrelated on clinical or other grounds to the other items in the factors onto which they loaded. These were the items 'thirsty', which loaded onto Factor 10 (Attention-seeking–comfort-seeking) and Factor 12 (Stoical–accepting), and 'tireless', which loaded onto Factor 7 (Enthusiastic about food–interested in food). 'Thirsty' and 'tireless' had the lowest communalities of any of the items. 'Sorrowful' loaded appropriately onto a number of factors, but also loaded onto Factor 7 (Enthusiastic about food–interested in food), within which factor it did not appear to be related to other items.

Each of the 12 factors identified in this way was considered to be interpretable as a domain of HRQL for a dog: vitality (Factor 1, Eager–keen), physical limitation (Factor 2, Stiff–sore), lethargy (Factor 3, Listless–reluctant), anxiety (Factor 4, Panicky–nervous), aggression (Factor 5, Aggressive–unresponsive), emotional upset (Factor 6, Whining–crying), appetite (Factor 7, Enthusiastic about food–interested in food), consistency of behaviour (Factor 8, At ease–consistent), mental disturbance (Factor 9, Confused–complaining), attention-seeking (Factor 10, Attention-seeking–comfort-seeking), sadness (Factor 11, Sorrowful–sad) and acceptance (Factor

12, Stoical–accepting). This range of HRQL domains was similar to that included in human HRQL instruments.

There was not a simple relationship between the factors and items loading significantly onto them, as shown in Table 5.3. Fewer than half of all items (48 items) loaded significantly onto only 1 factor. These were: eager, keen, outgoing, curious, bold, excitable, bright, nosy, alert, interested, fun-loving, stretching, stiff, sore, limping, uncomfortable, awkward, listless, reluctant, apathetic, sleepy, panicky, nervous, uneasy, frightened, upset, strained, cautious, aggressive, even-tempered, unresponsive, whining, crying, disturbed, enthusiastic about food, interested in food, greedy, tireless, consistent, inconsistent, obedient, at ease, confused, attention-seeking, comfort-seeking, clingy, stoical and accepting. Of the remaining items, a majority (49 items) loaded onto 2 factors, and a much smaller number of items loaded onto 3 factors (7 items) or 4 factors (5 items).

Table 5.4 shows the relationship between the behavioural domains in which disturbances were observed by owners of dogs suffering chronic pain, that were hypothesised to be relevant to the measurement of such pain, and the factors revealed by FA of responses obtained with an instrument developed from that hypothesis, each of which was interpreted as a domain of canine HRQL. The behavioural domains and associated descriptors appeared to have contributed to the measurement of various HRQL domains in an appropriate manner. For example, terms used to describe levels and types of activity were found to contribute to factors interpreted as HRQL domains relating to vitality, physical limitations, and lethargy, and those used to describe extroverted and introverted behaviour contributed to HRQL domains relating to vitality and to aggression

Table 5.3 Table of factors obtained through FA and the behavioural domains and associated items/descriptors loading onto each factor. Items loading only onto a single factor are printed in red, those loading onto 2 factors are in blue, those loading onto 3 factors are in green and those loading onto 4 factors are in plum.

Factor obtained from factor analysis	**Behavioural domains**	***A priori* positive and negative associated items/descriptors loading onto the factor** (*a priori* positive descriptors are printed in roman text and *a priori* negative descriptors are printed in italic)
Factor 1: Eager–keen	**Extroversion/introversion**	eager, outgoing, curious, bold, excitable, nosy, fun-loving, sociable, *quiet, subdued*
	Alertness	keen, inquisitive, bright, alert, interested
	Activity	energetic, lively, bouncy, boisterous, playful, active, *slowed, tired, lethargic, lacklustre, sluggish, weary*
	Comfort	stretching, comfortable
	Dependence	confident, independent
	Posture/mobility	athletic, fit, relaxed
	Contentment	contented, happy
	Anxiety	easy-going
Factor 2: Stiff–sore	**Posture/mobility**	*stiff, limping, awkward,* athletic, fit, relaxed
	Comfort	*sore, pained, uncomfortable,* comfortable
	Activity	*slowed, tired, apprehensive, weary,* energetic, lively, bouncy, boisterous, playful, active
	Contentment	*resigned, miserable, unhappy, sorrowful, sad,* contented, happy
	Dependence	*pathetic/pitiful*
	Agitation	*agitated, restless, unsettled*
	Anxiety	*distressed*

Table 5.3 Continued from page 170.

Factor obtained from factor analysis	Behavioural domains	*A priori* positive and negative associated items/descriptors loading onto the factor (*a priori* positive descriptors are printed in roman text and *a priori* negative descriptors are printed in italic)
Factor 3: Listless–reluctant	Activity	*slowed, lethargic, listless, lacklustre, reluctant, sluggish, apathetic, weary, sleepy*
	Contentment	*resigned, miserable, unhappy, sorrowful, sad*
	Alertness	*depressed, dull*
	Extroversion/introversion	*subdued, withdrawn, unsociable, detached*
	Dependence	*pathetic/pitiful*
	Alertness	*uninterested*
Factor 4: Panicky–nervous	Activity	*apprehensive*
	Anxiety	*panicky, nervous, uneasy, frightened, upset, strained, anxious, cautious, distressed*, easy-going, laid-back
	Agitation	*agitated, panting, restless*, calm
	Dependence	confident
Factor 5: Aggressive–unresponsive	Extroversion/introversion	*withdrawn, unsociable, unresponsive*, sociable, friendly, affectionate
	Aggression	*aggressive, irritable, grumpy, good*-natured, even-tempered, placid, territorial-protective
	Anxiety	easy-going
	Contentment	*unhappy*
	Agitation	*agitated, restless*
	Compulsion	*compulsive*
Factor 6: Whining–crying	Compulsion	*compulsive*
	Agitation	*whining, crying*
	Comfort	*moaning, groaning, complaining*
	Agitation	*panting, disturbed, restless, unsettled*
	Appetite	*picky (food), off his/her food*

Table 5.3 Continued from page 171.

Factor obtained from factor analysis	Behavioural domains	*A priori* positive and negative associated items/descriptors loading onto the factor (*a priori* positive descriptors are printed in roman text and *a priori* negative descriptors are printed in italic)
Factor 7: Enthusiastic about food–interested in food	Appetite	enthusiastic about food, interested in food, greedy, *picky (food), off his/her food*
	Contentment	*sorrowful*
	Activity	tireless
Factor 8: At ease–consistent	Anxiety	easy-going, laid-back, *anxious*
	Agitation	calm, at ease, *agitated*
	Consistency	consistent, *inconsistent*
	Extroversion/introversion	friendly, *quiet*
	Contentment	contented
	Aggression	good-natured, placid
	Alertness	obedient
Factor 9: Confused–complaining	Comfort	*complaining, pained, moaning, groaning*
	Contentment	*miserable*
	Alertness	*depressed, dull, confused*
	Aggression	*irritable, grumpy*
	Anxiety	*distressed*
Factor 10: Attention-seeking–comfort-seeking	Extroversion/introversion	*Detached*, affectionate
	Dependence	*attention-seeking, comfort-seeking, clingy*, independent
	Appetite	thirsty
Factor 11: Sorrowful–sad	Alertness	*Uninterested*, inquisitive
	Contentment	*unhappy, sorrowful, sad*, happy
	Aggression	territorial/protective
Factor 12: Stoical–accepting	Appetite	thirsty
	Anxiety	accepting
	Comfort	stoical

Table 5.4 Table showing how the behavioural domains and their items were related to the factors revealed by FA of ratings of those items, each of which factors was considered to represent a domain of HRQL. Factors included in the table are those with significant loadings (loadings >0.3) for at least one-third of the items of the relevant behavioural domain.

Behavioural domains	**Factors with significant loadings for at least one-third of the items of that behavioural domain**
Activity	Factor 1: Eager–keen Factor 2: Stiff–sore Factor 3: Listless–reluctant
Comfort	Factor 2: Stiff–sore Factor 6: Whining–crying Factor 9: Confused–complaining
Appetite	Factor 6: Whining–crying Factor 7: Enthusiastic about food –interested in food
Extroversion/introversion	Factor 1: Eager–keen Factor 5: Aggressive–unresponsive
Aggression	Factor 5: Aggressive–unresponsive
Anxiety	Factor 4: Panicky–nervous
Alertness	Factor 1: Eager–keen
Dependence	Factor 1: Eager–keen Factor 10: Attention-seeking –comfort-seeking
Contentment	Factor 2: Stiff–sore Factor 3: Listless–reluctant Factor 11: Sorrowful–sad
Consistency	Factor 8: At ease–consistent
Agitation	Factor 2: Stiff–sore Factor 4: Panicky–nervous Factor 6: Whining–crying Factor 8: At ease–consistent
Posture/mobility	Factor 1: Eager–keen Factor 2: Stiff–sore
Compulsion	Factor 5: Aggressive–unresponsive Factor 6: Whining–crying

Since the FA of GUVQuest item responses revealed an underlying factor structure that compared well with the hypothetical construct upon which instrument was developed – that of canine chronic pain as an unpleasant sensory and emotional experience that has an impact upon a range of domains of HRQL (and is measurable by its impact upon those domains) – it was considered to have provided some evidence for the construct validity of the GUVQuest.

5.2.2 Calculating scores

5.2.2.1 Methods

On the basis of the chosen 12-factor model, in which each factor was considered to represent a domain of HRQL, a score for each HRQL domain was calculated, providing an HRQL 'profile' for each dog at each assessment. Each HRQL domain score was obtained by calculating the mean of all item ratings for the relevant factor. However, 8 factors included ratings for both positively- and negatively-loading descriptors, with high scores for one type of descriptor being accompanied in most cases by low scores on the other type. A simple arithmetical operation would not take account of this, and calculating the mean of the raw scores would become meaningless. Consequently, reversals to ratings were made for those items loading with the opposite sign (+ or -) to the factor's principal items. Thus, in the case of Factor 1 (Eager–keen), scores for descriptors loading with the opposite sign to the principal items for that factor (e.g. scores for 'quiet', 'slowed', etc.) were reversed, so that a score of 1 became a score of 5 and a 5 became 1, a score of 2 became a score of 4, one of 6 became one of 0, and so on, as used by Hsu and Serpell when developing a questionnaire for measuring temperament in pet dogs (Hsu and Serpell, 2003; Hsu, personal communication).

5.2.2.2 Results

The HRQL domain scores for each of the dogs (with >1 questionnaire) in the Hospital (DJD), Practice and Control groups are given in Appendix 19. These

HRQL domains retain the numbering of the factors by which they were identified, e.g. factor 1 (Eager–keen) becomes HRQL domain 1 (Eager–keen).

5.2.3 Testing Hypothesis 2

HRQL domain scores calculated for dogs in the Practice and Hospital (DJD) groups and for dogs in the Control group for which >1 questionnaire was completed, were used to test whether the GUVQuest was able to discriminate between groups known to differ on the attribute of interest: chronic pain.

5.2.3.1 Methods

Using MINITAB for Windows® (Release 13), discriminant analysis with cross-validation was carried out on the sets of HRQL domain scores obtained from 250 questionnaires (all questionnaires for dogs in Control group, in Hospital (DJD) group and in Practice group), of which 32 had been completed for Control group dogs with clinician pain scores of 0, and the remainder were for dogs in the Practice and Hospital (DJD) groups, with a range of clinician pain scores. Since a small number (7) of dogs in the latter group were awarded pain scores of 0 at the end of treatment, the analysis was subsequently repeated with those dogs included with the Control group dogs in a 'no pain' group.

5.2.3.2 Results

The HRQL domain scores were able to discriminate between dogs in a Chronic pain group (Hospital (DJD) and Practice groups combined) and those in the Control group in 86% of cases, with >93% of questionnaires completed for Control group dogs and >84% of those completed for Chronic pain group dogs being correctly categorised, as shown in Table 5.5.

An examination of the misclassified questionnaires revealed that 6 of these were where a Chronic pain group dog had been awarded a pain score of 0 (at end of treatment). A further 9 misclassified questionnaires had no clinician pain score, but 8 of these were the last questionnaires completed for that dog, when pain scores may have been 0 or close to 0 if treatment was at an end.

Table 5.5 Results of discriminant analysis performed with cross-validation on HRQL domain scores (n questionnaires=250) that were required to discriminate between dogs in Control group (Control group) and those in Hospital (DJD) and Practice groups (together forming Chronic pain group).

	Controls group	**Chronic pain group**
N identified as belonging to controls group	30	33
N identified as belonging to chronic pain group	2	185
N total	32	218
N correctly identified	30	184
Proportion correctly identified	0.938	0.849

Consequently, it may be that that these questionnaires were originally misclassified as questionnaires completed for dogs with chronic pain.

If this is the case, then the number of false negative results is reduced to 18, representing only 8% of questionnaires considered. Examination of individual misclassifications further reveals that 10 of these misclassified questionnaires were completed by only 2 owners, who may have had difficulty completing the questionnaire. If all of the questionnaires completed by those 2 owners were also excluded from the analysis, then the number of inexplicable misclassifications is reduced to 8 questionnaires out of a total of 204. This represents a misclassification (false negative) of <4% of questionnaires.

When dogs in the Practice and Hospital (DJD) groups with pain scores of 0 were included, with the Control group dogs, in a group titled 'no pain', the HRQL domain scores were able to discriminate correctly between the 'no pain' group and a 'some pain' group (dogs with pain scores of 1 or more) in >88% of all cases for which pain scores were available (n=118), with nearly 95% of all 'no pain' dogs and >86% of 'some pain' dogs being correctly categorized, as shown in Table 5.6.

An examination of those questionnaires that were misclassified in this second analysis revealed that 11 of the 23 misclassifications were questionnaires completed by just 2 owners. A further 8 questionnaires were misclassified in both first and second analyses: these questionnaire are deserving of a closer look, since it may be that those respondents were having difficulty in completing the questionnaire, or were not taking care to complete it correctly. If all questionnaires completed by the two owners previously mentioned are excluded, along with those 8 questionnaires that were misclassified in both analyses, then only 4 questionnaires out of 157 (<3%) were incorrectly categorised (according to clinicians' pain scores).

These results provided some evidence for the construct validity of the GUVQuest, since the scores it generated were able to discriminate well

Table 5.6 Results of discriminant analysis performed with cross-validation on HRQL domain scores from questionnaires for which a clinician pain score was available (n=218) in which scores were required to discriminate between dogs awarded a clinician pain score of 1 or more (classified as 'some pain') and dogs awarded a clinician pain score of 0 (classified as 'no pain').

	Cases classified as 'no pain'	**Cases classified as 'some pain'**
N identified as belonging to 'no pain' group	37	24
N identified as belonging to 'some pain' group	2	155
N total	39	179
N correctly identified	37	155
Proportion correctly identified	0.949	0.866

between groups known to differ on the attribute the instrument was intended to measure.

5.2.4 Evaluating Hypothesis 3

A comparison of initial HRQL domain scores obtained for Control group dogs and dogs in the Hospital (DJD) group was expected to reveal clear differences between the ranges of scores for the two groups. Subsequently, over time, the HRQL domain scores obtained for the Control group dogs were expected to change little. Conversely, scores for dogs in the Hospital (DJD) group, most of which would be expected to demonstrate clinical improvement over the treatment period, were expected to change over time to more closely resemble the scores obtained for the Control group dogs (with the scores increasing or decreasing depending upon the HRQL domain concerned).

5.2.4.1 Methods

In order to identify a range of HRQL domain scores obtained for healthy dogs and the stability of these over time, the scores for those Control group dogs for which two separate questionnaires were completed were examined. Summary statistics were used to reveal the range of HRQL domain scores for this group on two occasions: at time of completing the initial questionnaire and at time of completing the follow-up questionnaire. Graphs of the resulting mean, median, minimum and maximum scores were plotted for each HRQL domain, with graphs for initial questionnaire and follow-up questionnaire overlaid, to examine the stability of HRQL domain scores over time for this group.

To compare these results at the group level with HRQL domain scores for dogs with DJD, and the way in which their scores may change during treatment, similar summary statistics were calculated and similar graphs were prepared for the first and last questionnaires completed by owners of all dogs in the Hospital (DJD) group.

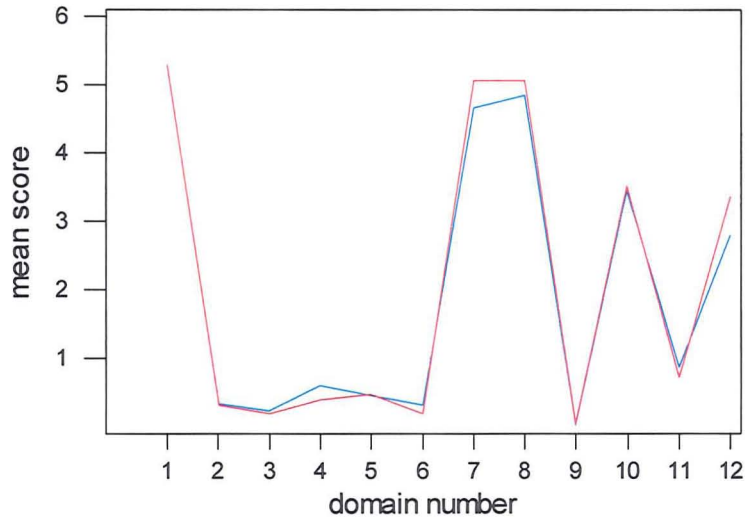
5.2.4.2 Results

The mean, median, minimum and maximum HRQL domain scores for dogs in the Control group (n dogs=16), for initial questionnaire and for follow-up questionnaire, are given in Appendix 20, and are illustrated in Figures 5.2(a), 5.3(a), 5.4(a) and 5.5(a). These summary statistics and their graphical representations revealed that scores for most HRQL domains tended to be relatively stable over time for Control group dogs. The profile of HRQL domain scores illustrated graphically shows that these dogs tended to receive high scores for HRQL domains 1, 7 and 8 (associated, respectively, with vitality, appetite and consistency of behaviour), and low scores for HRQL domains 2-6 (associated, respectively, with physical limitations, lethargy, anxiety, aggression, and emotional upset), with scores for HRQL domains 10-12 (attention-seeking, sadness, and acceptance) spread in the middle range. The scores generated by the Control group dogs may be considered for the purposes of this study to represent scores for a healthy population, with which to compare the scores for dogs with chronic pain.

The mean, median, minimum and maximum HRQL domain scores for dogs in the Hospital (DJD) group (n cases=61), for the initial questionnaire and for the last questionnaire completed for each dog, are given in Appendix 21 and are illustrated in Figures 5.2(b), 5.3(b), 5.4(b) and 5.5(b). At the group level, these scores appear to be different from those obtained from the Control group. At the beginning of treatment, the dogs with DJD had scores for HRQL domain 1 (associated with vitality) that were considerably lower, and those for HRQL domains 7 (appetite) and 8 (consistency of behaviour) that were somewhat lower, than the scores obtained for Control group dogs. Dogs with DJD also had scores that were considerably higher than the Control group for HRQL domains 2, 3, 4, 5, and 6 (physical limitation, lethargy, anxiety, aggression, and emotional upset), and slightly higher scores for domains 10-12 (attention-seeking behaviour, sadness, and acceptance).

Figure 5.2 Graph of mean score plotted against HRQL domain number for **(a)** Control group and **(b)** Hospital (DJD) group dogs. Graphs for scores generated by initial questionnaire (blue line) and follow-up questionnaire (red line) are overlaid on the same graph.

(a)



(b)

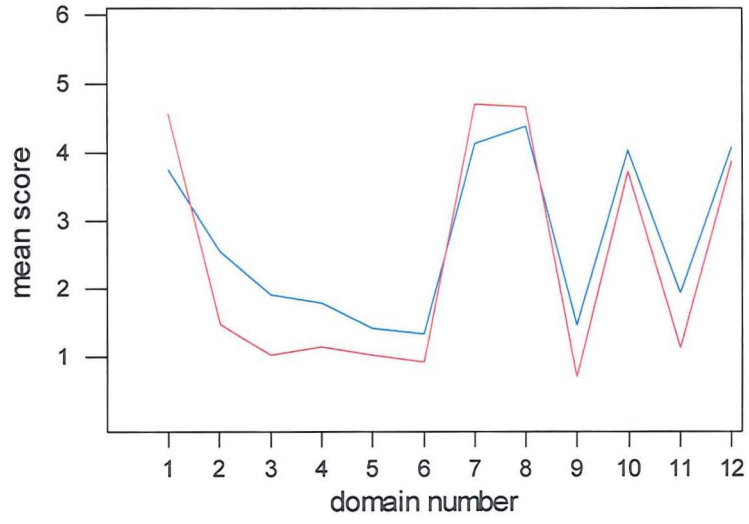
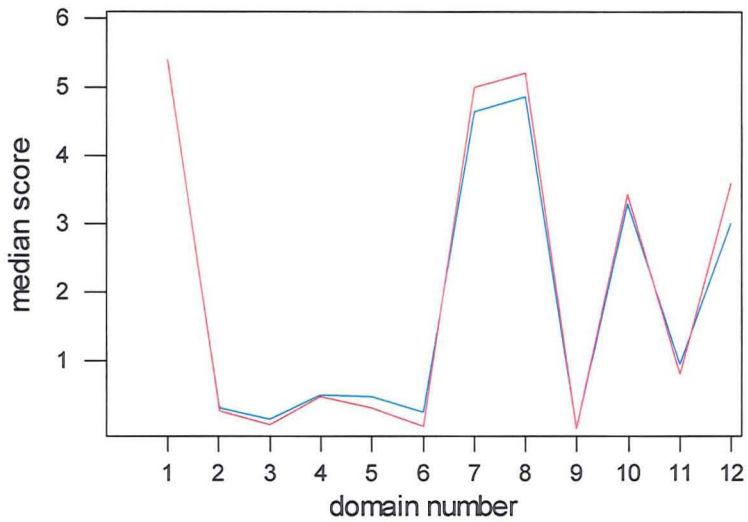


Figure 5.3 Graph of median score plotted against HRQL domain number for **(a)** Control group and **(b)** Hospital (DJD) group dogs. Graphs for scores generated by initial questionnaire (blue line) and follow-up questionnaire (red line) are overlaid on the same graph.

(a)



(b)

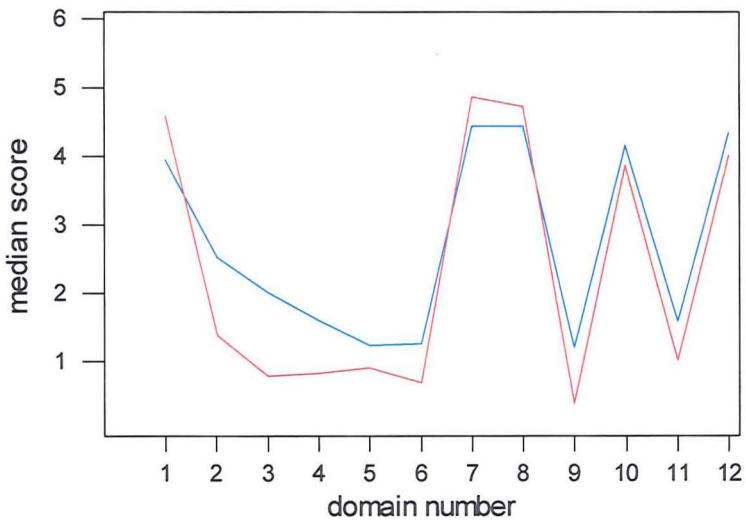
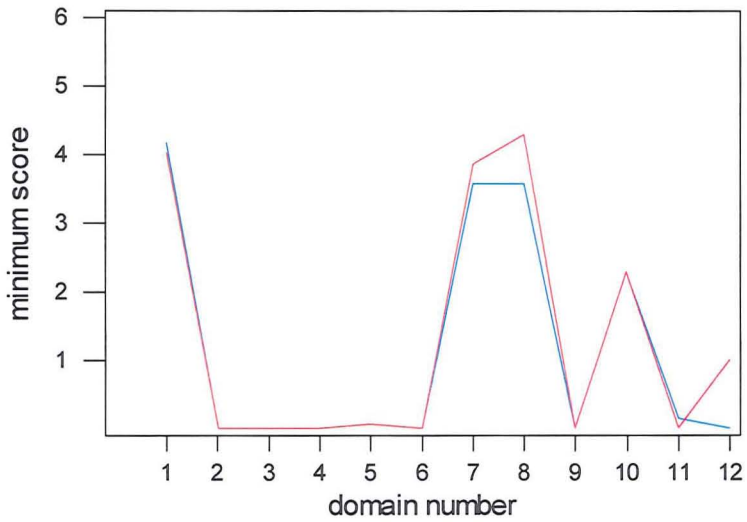


Figure 5.4 Graph of minimum score plotted against HRQL domain number for **(a)** Control group and **(b)** Hospital (DJD) group dogs. Graphs for scores generated by initial questionnaire (blue line) and follow-up questionnaire (red line) are overlaid on the same graph.

(a)



(b)

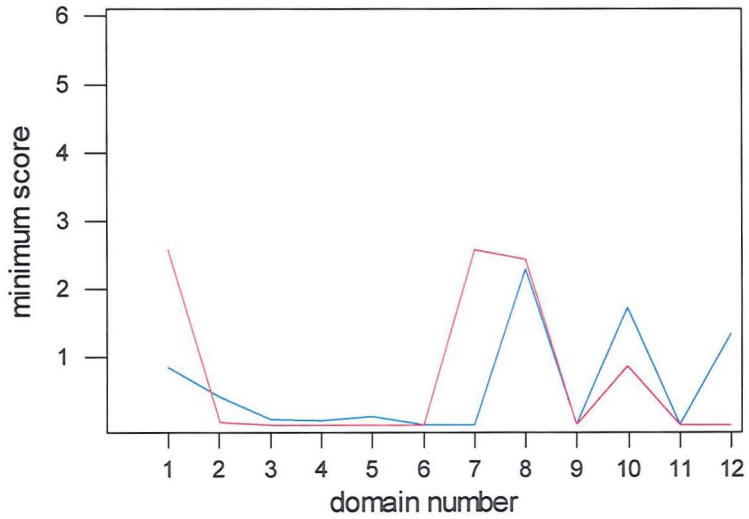
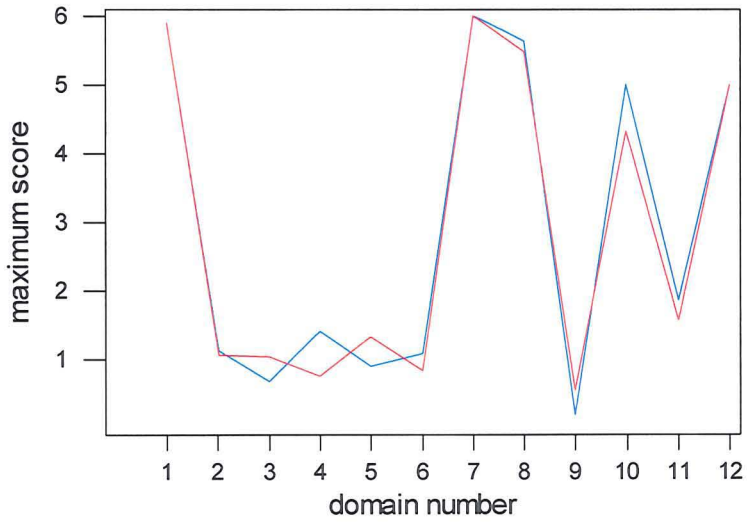
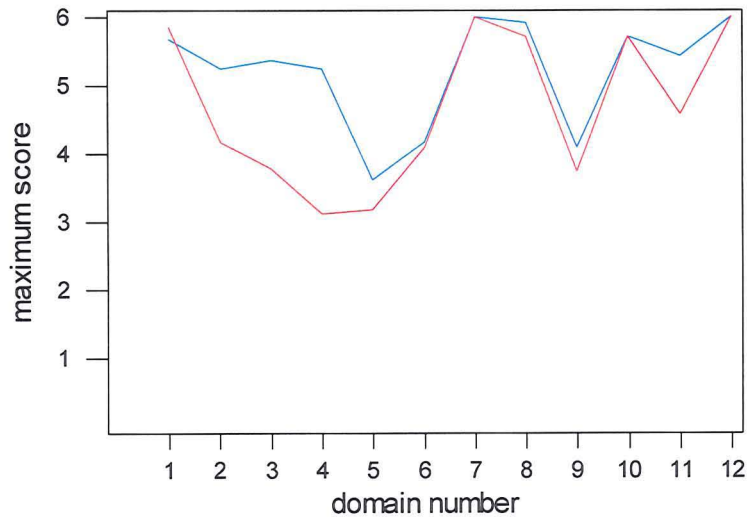


Figure 5.5 Graph of maximum score plotted against HRQL domain number for **(a)** Control group and **(b)** Hospital (DJD) group dogs. Graphs for scores generated by initial questionnaire (blue line) and follow-up questionnaire (red line) are overlaid on the same graph.

(a)



(b)



However, a comparison of HRQL domain scores obtained from the first questionnaires (indicated by the blue line), with those obtained from the last questionnaires (indicated by the red line), as shown on Figures 5.2(b) and 5.3(b), revealed that over time, with treatment, the scores for domain 1 and, to a lesser degree, domains 7 and 8 showed a tendency to increase, and those for most other domains showed a tendency to decrease, so that in general the HRQL domain scores obtained from the last questionnaires completed for dogs with DJD were closer to the scores obtained for the Control group dogs, as was predicted. This is illustrated in Figure 5.6, where graphs of mean domain scores for Hospital (DJD) group dogs are overlaid with those for Control group dogs, for initial questionnaires (graph a) and last questionnaires (graph b) completed for each group. It can be seen that the HRQL profiles for the last questionnaires completed for the two groups of dogs were closer to each other than were the profiles for the initial questionnaires completed.

5.2.5 Evaluating Hypothesis 4

This evaluation was intended to obtain evidence for the validity of an instrument that was designed to evaluate change in individual cases, using individuals as their own controls, by examining how changes in HRQL domain scores related to clinical change for individual dogs.

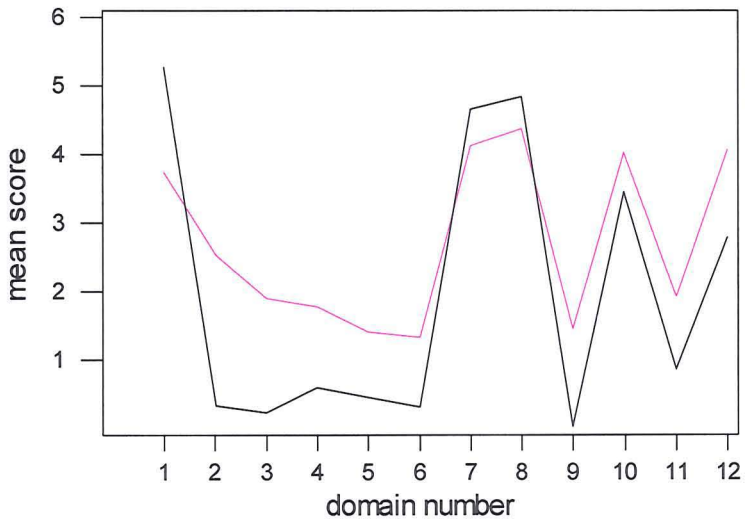
5.2.5.1 Methods

The HRQL domain scores for individual cases were plotted against questionnaire number. As each questionnaire was completed at the time of a hospital or practice consultation, this meant that the scores were plotted over time and, for non-Control group dogs, over a treatment period.

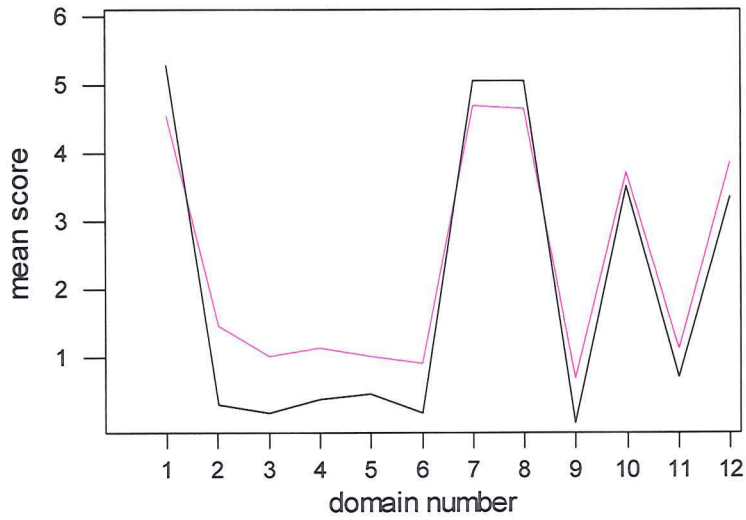
An examination was made of changes in HRQL domain scores over time, and the extent to which these reflected clinical change. The evidence for clinical change was derived from several different indices included in clinician and owner questionnaires, for which some evidence for validity as measures of clinical change had been obtained, as reported in Chapter 4.

Figure 5.6 Graph of mean score plotted against HRQL domain number for **(a)** initial questionnaires completed, and **(b)** last questionnaires completed, for Hospital (DJD) group (magenta line) and Control group (black line) dogs (overlaid on same graph).

(a)



(b)



5.2.5.1.1 Selection of individual cases

A selection of cases was made to represent a range of ages and breeds and a variety of patterns of clinical change in both hospital and practice settings. These were chosen from all cases for which the same clinician examined the dog on each visit to hospital or practice, and completed a clinician questionnaire on the day of the consultation, and for which the same owner completed consecutive questionnaires, each on the correct day. A selection of dogs was made from the Control group, using the same criteria.

5.2.5.1.2 Indices of clinical change with which comparison of HRQL domain scores was made

The assessment of clinical change was based upon a range of indicators, for each of which some evidence for its validity had been obtained (see Chapter 4, pages 145-148). These were the clinician's pain scores and estimates of change from one examination to the next, and the owner's responses to transition questions about global change in 8 behavioural domains: activity, pain, sociability, aggression, anxiety, enthusiasm, happiness and mobility.

5.2.5.1.3 Presentation of results

A graph of HRQL domain scores over time for each of the selected Control group dogs was accompanied by a brief description of the dog. A similar graph for each of the cases selected from the Hospital (DJD) and Practice groups was also accompanied by that information and, in addition, by a summary of clinician and owner ratings, and resulting conclusions about clinical change inferred from those indices, followed by an interpretation of the extent to which evidence for clinical change was provided by changes in HRQL domain scores.

5.2.5.2 Results

The graphs of HRQL domain scores against questionnaire number for 13 selected cases were plotted: 4 Control group dogs, 3 Practice group dogs and 6 dogs from the Hospital (DJD) group. These graphs, and accompanying

information about each of the dogs, indices of clinical change and interpretation of the extent to which clinical change was reflected in changes in HRQL domain scores, are shown in the graphs and given in the accompanying text in Appendix 22. One example from each group is given in Figure 5.7 (Control group, Dog D), Figure 5.8 (Hospital (DJD) group, Dog F) and Figure 5.9 (Practice group, Dog G).

Example of Control group dog

The first example, Dog D, was a female entire Border Collie belonging to the Control group. She was 4 years old at first consultation. Her second examination was carried out (and second questionnaire completed) 64 days after the first. Graphs of HRQL domain scores from initial and follow-up questionnaires are shown in Figure 5.7. These show little change over time for HRQL domains 1-10, with some change apparent for domains 11 and 12.

Example of Hospital (DJD) group dog

Dog F belonged to the Hospital (DJD) group. He was a male, entire Collie, 10 years old at first consultation. At first consultation, the owner reported that the dog had been in pain for 3 weeks. Dog F was treated surgically for DJD. The relevant clinician ratings and owner transition ratings are as follows:

Clinician pain scores and assessments of change

Consult	Day	Pain score	Acute exacerbation? (baseline)	Change assess.
1	0	6	No	
2	42	3	No	I

Owner transition question ratings

Consult	Activity	Pain	Sociab.	Aggress.	Anxiety	Enthus'm	Happiness	Mobil'y
1	GD	GI	NC	NC	SI	D	NC	GD
2	I	GD	NC	NC	NC	I	NC	I

From these ratings it was inferred that there had been an improvement in Dog F's clinical condition at the second consultation compared with the first. This evidence for clinical improvement was reflected in the changing HRQL domain scores for this dog, which are illustrated in Figure 5.8.

Figure 5.7 Graphs of HRQL domain scores plotted against questionnaire number for Dog D, an example of a Control group dog.

Key: Domains 1 and 7 – solid black line
Domains 2 and 8 – solid red line
Domains 3 and 9 – solid blue line
Domains 4 and 10 – dashed black line
Domains 5 and 11 – dashed red line
Domains 6 and 12 – dashed blue line

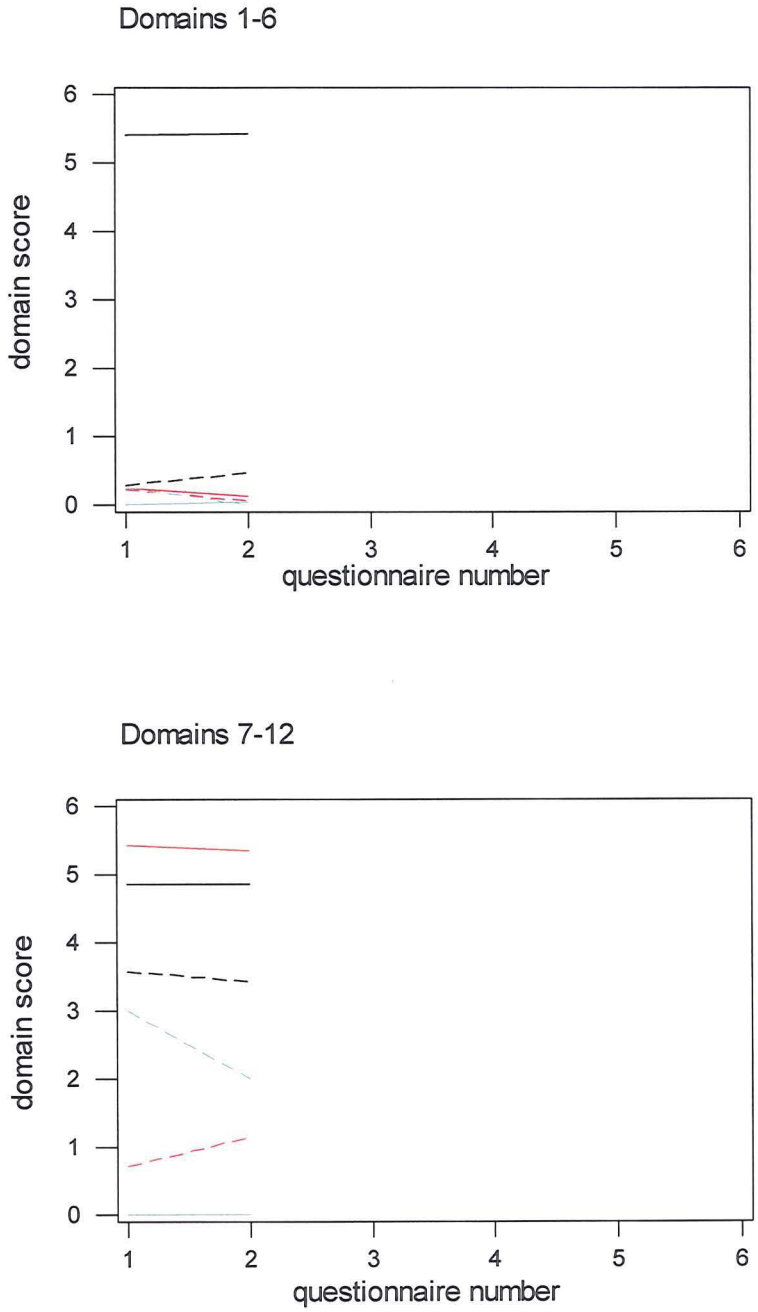
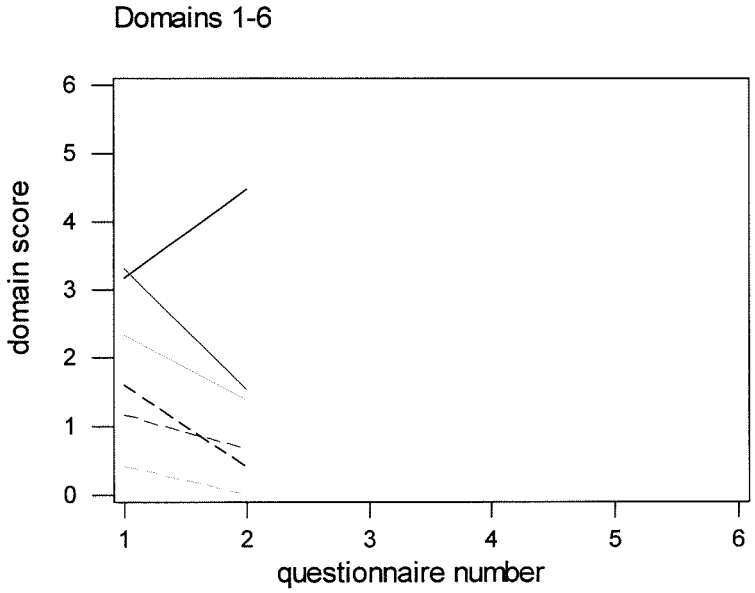
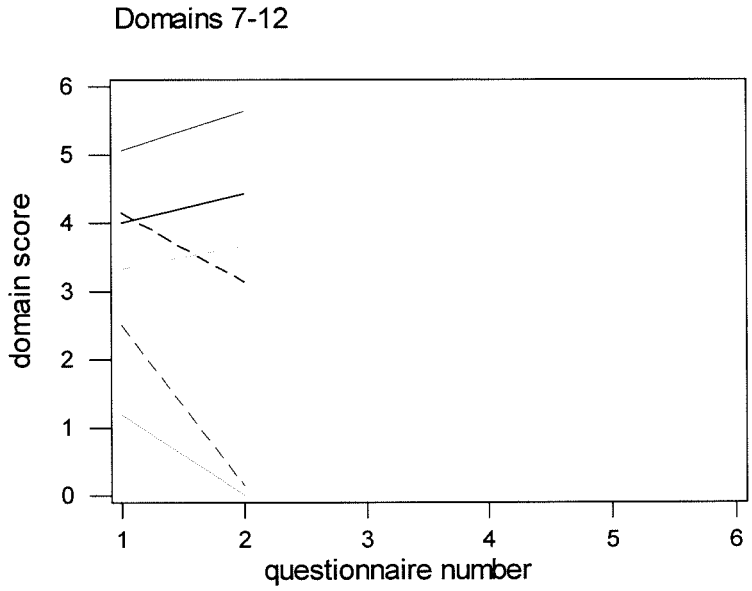


Figure 5.8 Graphs of HRQL domain scores plotted against questionnaire number for dog F, an example of a Hospital (DJD) group dog.

Key: Domains 1 and 7 – solid black line
Domains 2 and 8 – solid red line
Domains 3 and 9 – solid blue line
Domains 4 and 10 – dashed black line
Domains 5 and 11 – dashed red line
Domains 6 and 12 – dashed blue line





Example of Practice group dog

Dog G belonged to the Practice group. She was a female, neutered Labrador retriever, 10 years 10 months old at her first consultation, at which her owner reported that she had been in pain for 1 month. Dog G was treated conservatively for DJD. The relevant clinician ratings and owner transition ratings are as follows:

Clinician pain scores and assessments of change

Consult	Day	Pain score	Acute exacerbation? (baseline)	Change assess.
1	0	7	No	
2	7	6	No	SI
3	14	6	No	NC
4	21	5	No	SI
5	35	4	No	I
6	49	3	No	I
7	70	2	No	SI

Owner transition question ratings

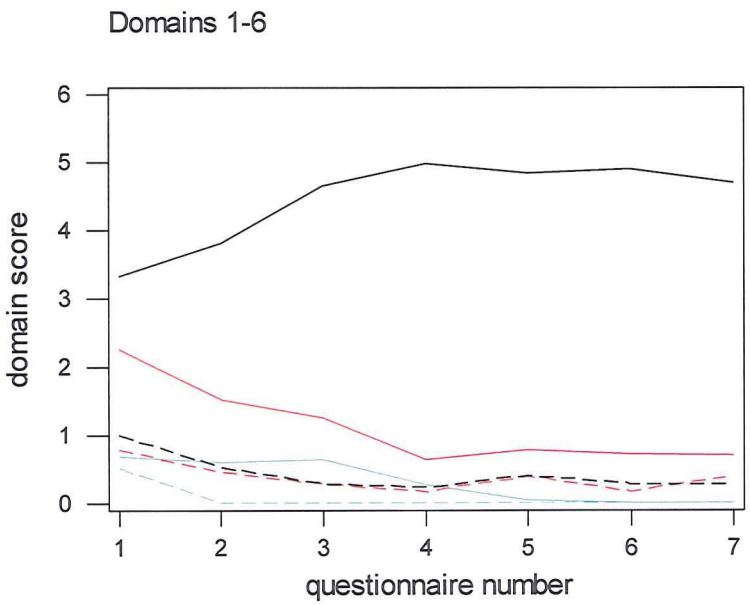
Consult	Activity	Pain	Sociab'y	Aggress'n	Anx'y	Enthus'm	Happ's	Mobil'y
1	D	I	NC	NC	NC	SD	SD	D
2	I	D	SI	NC	NC	SI	SI	I
3	I	D	I	NC	NC	SI	SI	I
4	I	D	I	*	*	*	*	*
5	I	D	SI	NC	NC	SI	I	I
6	I	D	I	NC	NC	I	I	I
7	NC	NC	NC	NC	NC	NC	NC	NC

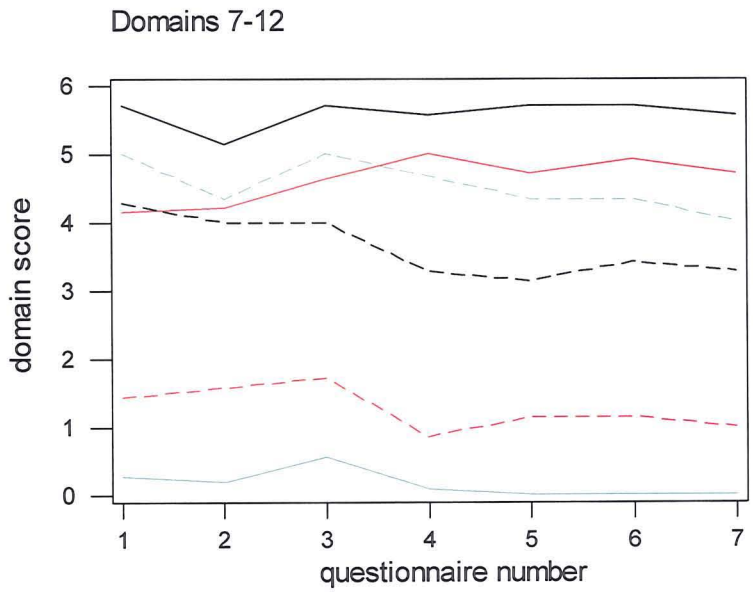
From these ratings it was inferred that there had been some degree of improvement in clinical condition throughout treatment. This evidence for clinical change was reflected in the HRQL domain scores obtained for this dog over the treatment period, which are shown in Figure 5.9.

An examination of the HRQL domain scores for dogs A to M (Appendix 22), and the ways in which these changed for individual dogs receiving treatment for a chronic and painful condition, revealed predictable profiles and predictable patterns of change during the course of treatment, in accordance with other evidence for clinical change in individual dogs.

Figure 5.9 Graphs of HRQL domain scores plotted against questionnaire number, for dog G, an example of a Practice group dog.

Key: Domains 1 and 7 – solid black line
Domains 2 and 8 – solid red line
Domains 3 and 9 – solid blue line
Domains 4 and 10 – dashed black line
Domains 5 and 11 – dashed red line
Domains 6 and 12 – dashed blue line





Although the patterns of HRQL domain scores differed between individuals, any changes in scores tended to reflect clinical change (whether this was assessed to be an improvement or deterioration in condition). The evidence that HRQL domain scores for individual dogs reflected clinical change (inferred from both clinician and owner reports) supported Hypothesis 4 and provided evidence for the validity of the GUVQuest for the evaluation of change in individual cases, using individuals as their own controls.

5.3 Discussion

The tests and evaluations described in this chapter provided some evidence for the construct validity of the instrument developed in the course of this study.

Multivariate FA revealed an interpretable 12-factor model that comprised a range of HRQL domains that were relevant to the dog and were similar to those included in HRQL instruments designed to measure human chronic pain, thereby supporting Hypothesis 1.

This 12-factor model accounted for over 65% of the variance in the data set from which it was created. The % variance accounted for by a factor model is a measure of its usefulness, with higher figures representing better models. The figure obtained for the GUVQuest approximates that reported for many self-report human pain and HRQL instruments, and exceeds that reported for a number of proxy human pain and HRQL instruments, and dog temperament instruments, such as those reported in the introduction to this chapter.

In the matrix of behavioural domains and descriptors considered relevant to measuring chronic pain in dogs, each item was associated with only one behavioural domain. Conversely, FA of item ratings revealed that fewer than half of all items (48 items) loaded only onto one factor. The remaining items loaded onto 2 factors (49 items), 3 factors (7 items) or 4 factors (5 items). An

examination of those items with multiple loadings revealed that most of these loadings were sensible. For example, the item 'slowed' contributed to Factor 1 (Eager–keen), Factor 2 (Stiff–sore) and Factor 3 (Listless–reluctant), and such a descriptor might be expected to contribute to those factors, which related to vitality, physical limitation and lethargy, respectively. Similarly, the item 'good-natured' loaded appropriately onto Factor 5 (Aggressive–unresponsive) and Factor 8 (At ease–consistent), as did the item 'irritable'. The item 'unhappy' was one that loaded onto 4 factors, Factor 2 (Stiff–sore), Factor 3 (Listless–reluctant), Factor 5 (Aggressive–unresponsive) and Factor 11 (Sorrowful–sad), each of which association was considered to be sensible. This lack of a 'simple structure', in which the variables only load substantially onto a single factor, is not unusual, and has been noted in a recent study of the factor structure of the MPQ (Coste *et al.*, 2005), the items of which, like the core items of the GUVQuest, consist of simple, single-word terms.

It is important to recognise that the behavioural domains of the matrix upon which the instrument was constructed represented domains of behaviour in which owners had reported disturbances, and were validated as such (as domains of behaviour, not as domains of HRQL). Nevertheless, an examination of the relationship between behavioural domains and associated descriptors hypothesised to be relevant to the measurement of a dog's chronic pain, and the domains of HRQL revealed by FA of owner ratings of those descriptors, revealed that most descriptors appeared to be contributing in an appropriate manner to the measurement of a range of HRQL domains. For example, most of the items selected to describe the behavioural domain 'Activity' were found to contribute to HRQL domains (load on to the relevant factors) associated with vitality (Factor 1: Eager–keen), physical limitation (Factor 2: Stiff–sore) and lethargy (Factor 3: Listless–reluctant). In some cases, the relationship was simpler, with all of the items chosen to describe the behavioural domain 'Aggression' loading significantly onto HRQL domain 5 (Aggressive–unresponsive), and almost all of those for the

behavioural domain 'Anxiety' loading significantly onto HRQL domain 4 (Panicky–nervous).

The factor structure underlying the GUVQuest's item responses was considered to compare well with the hypothetical construct upon which the instrument was developed. The range of factors (the nature of each one identified by the descriptors loading and not loading onto it) was similar to the range of domains included in human HRQL instruments, including vitality, physical limitation, lethargy, anxiety, aggression, emotional upset, appetite, dependence (attention-seeking), sadness and acceptance. The nature of some of the other factors reflected domains considered relevant to measuring a dog's chronic pain that were identified through interviews with dog owners, including reports of 'good days and bad days' and mental disturbance. Consequently, the factor model revealed by FA, including associated item loadings, was considered to represent a range of canine HRQL domains that are affected by chronic pain, and so provided some evidence for the construct validity of the GUVQuest.

Initially, a data set for FA was considered that included all of the cases within the Practice group and all of the DJD cases within the Hospital group (including cases with only one questionnaire), in the hope that this data set would be sufficiently large to split into two in order to carry out exploratory factor analysis on one half of the data set, followed by confirmatory factor analysis on the other half. Unfortunately, the number of items of missing data meant that each of these separate data sets was not sufficiently large for factor analysis to be carried out, so that only one FA could be undertaken.

As the instrument was intended primarily to be an evaluative tool, rather than a discriminative one, it was decided to perform the FA on data obtained for dogs that went on to complete more than one questionnaire. The dataset used for this analysis therefore comprised the item ratings obtained for all dogs with DJD in Hospital and Practice groups for which more than one

questionnaire had been completed. Consideration was given to whether or not also to include the data for dogs within the Control group for which more than one questionnaire had been completed; ratings on many of the items for this group tended to be at the ends of the scale (0/1 or 5/6) with the potential to enhance any result. Graphs of eigenvalues against principal components were plotted using both of the proposed datasets, and it could be seen from the similarities between these scree plots that including or excluding the Control group dogs from the FA made little impact upon the amount of variability explained. To ensure that the factor structure did not differ significantly if the data for Control group dogs was included or excluded, FA was carried out on each data set. The 12-factor models were not identical for each data set, but there was a significant overlap between them, with 9 of the factors being common to both analyses, containing very similar collections of items, and accounting for over 50% of the variance in their respective models. (The items contained in the remaining 3 factors in each model were associated with unhappiness and lethargy in each case.) Consequently, it was considered that the decision to base subsequent analysis on the factor model obtained from FA of the ratings provided for dogs in the Hospital (DJD) and Practice groups only was an appropriate one.

Determining the correct number of components to retain in FA is a crucial step in this kind of instrument development, and it is recommended that a 'careful and diversified approach' to this question is taken, including the use of more than 1 rule of extraction (as used in this study, although other rules of extraction may be more suitable than those used) and comprehensively assessing the models obtained (which assessment has begun to be undertaken in this study) (Coste *et al.*, 2005). Coste and colleagues (2005) also recommend repeating the analysis across samples (insufficient cases made this impossible in this study) and considering complimentary methods of confirmatory factor analysis. It is recognised that such an approach during future field-testing of the instrument may result in the selection of a different factor model upon which to base refined versions of the instrument.

However, based upon the data and methods of analysis available in this study, a 12-factor model was established as the most appropriate, and the identification of its 12 factors as domains of HRQL led to the calculation of a score for each HRQL domain. Such use of FA to reveal separately scored domains has been recommended (Streiner and Norman, 1995), and has been adopted by other proxy instrument developers. For example, Armstrong and colleagues, when developing the Miami Pediatric Quality of Life Questionnaire–Parent Scale (Armstrong *et al.*, 1999) used factor analysis to identify a 3-factor structure. A score for each of these factors – social competence, emotional stability and self-competence – (and an aggregated total score for all three) was then used to explore the ability of the instrument to discriminate between known groups.

Examples of human HRQL instruments that provide separate scores for a number of domains or dimensions include the SF-36, which generates scores for 8 different domains representing 'health status concepts' (Ware and Sherbourne, 1992): physical functioning, social functioning, role limitations (physical and, separately, emotional), social functioning, bodily pain, general mental health, vitality, and general health perceptions. These scores can be examined separately for impact of health state or treatment effect, or can be summed to provide scores for physical health and mental health. The proxy HRQL instrument, the Royal Marsden Hospital Paediatric Quality of Life Questionnaire (Watson *et al.*, 1999) was designed to generate separate scores for functional status, global quality of life, physical symptoms, emotional status, social functioning, cognitive functioning, behavioural problems and progress at school. The PedsQL™, another paediatric QoL instrument, contains 4 separate generic core scales (physical, emotional, social and school functioning) and separate disease-specific modules, such as the PedsQL™ 3.0 Rheumatology module containing scales for pain and hurt, daily activities, treatment, worry and communication (Varni *et al.*, 2002a).

Each of these human self-report and proxy instruments, therefore, provides a profile of scores for a range of relevant HRQL domains. Such an approach to scoring can be important both in the development of the instrument and in its clinical application. In clinical use, certain HRQL domains may be more sensitive to clinical status or treatment effects than others, and during instrument development convergent validity can be established by comparing certain domain scores with appropriate established measures.

For this reason, an instrument's ability to provide a score for each separate domain of HRQL may be considered to be a methodological strength, and future refinement of the GUVQuest should consider methods of ensuring that the separately scored HRQL domains are discrete. Thus, a first step might be to remove from the instrument some of those items that were found to contribute to 3 or 4 different HRQL domains, particularly where the loadings on those domains were low. For example, the items 'unhappy' and 'restless' each loaded onto 4 domains, in all cases with loadings of less than 0.4. However, before any items were removed from the instrument it would be advisable to carry out additional field-testing with dogs suffering chronic orthopaedic pain, and with dogs suffering chronic pain of other causes, since the results of the FA carried out in this study must be considered to be tentative until confirmed by further field-testing and analysis.

Hypothesis 2 was supported when HRQL domain scores based on the chosen 12-factor model were found to discriminate well between dogs suffering chronic pain and dogs that were free from chronic pain, achieving figures for correct discrimination that were at least similar to those reported for the PICIC, a proxy instrument for pain measurement in communicatively impaired children, which correctly classified 87% of pain and non-pain episodes (as identified by caregivers) and was considered by its developers to have 'reasonable' ability to distinguish between such episodes (Stallard *et al.*, 2002). This instrument consisted of a checklist of 6 'core cues' used by caregivers as signs of definite or severe pain in their child (crying

with/without tears; screaming, yelling, groaning or moaning; screwed up or distressed looking face; body appears stiff or tense; difficult to comfort or console; flinches or moves away if touched), and pain or non-pain episodes were classified by the caregivers themselves.

This discriminative ability of the GUVQuest may be useful, since, although the instrument was intended primarily for evaluative purposes, it could also be used to alert the clinician to the possible presence of chronic pain when this may not be readily apparent because the behaviour changes associated with chronic pain tend to be subtle. Because of the risk of false negative results, in clinical use the GUVQuest could not be used to rule out the presence of chronic pain. However, it might be used as an alarm signal for the possible presence of prolonged pain, as has been proposed by Gauvain-Piquard and colleagues (1999) for the DEGR®. Such a purpose would be appropriate for the GUVQuest, provided that clinicians were aware of the risk of false positive and, more importantly, false negative results, but such risks would require to be quantified before the GUVQuest could be used in this way.

The HRQL domain scores obtained for Control group dogs, and the stability of these over time, may be considered to represent HRQL domain scores for a healthy dog. When these scores were compared at the group level with the range and stability of HRQL domain scores for dogs suffering chronic pain, it was found that there were clear differences between the ranges of scores obtained for these two groups and between the stability of scores over time for the two groups, which supported Hypothesis 3. However, the small size of the Control group means that the scores obtained for this sample may not reflect the true variability of the healthy population, and any conclusions reached on the basis of these scores must be considered to be tentative at this stage. Further field-testing of the instrument should include a larger healthy Control group in order to extend the results of this study.

The next steps in the development of the GUVQuest should include the exploration of instrument refinement by various means. One way in which instrument developers seek to refine an instrument is to exclude any items that contribute little to the measurement of interest, and often the communality of an item (how much of the variability of an item is accounted for by the underlying factor model) can be a useful guide in this regard. Low communalities are not interpreted as evidence that the data fail to fit the hypothesis, but simply that the variables analysed have little in common with each other (Darlington, n.d.). However, communalities must be interpreted in relation to the interpretability of the factors. A communality of 0.75 seems high but is meaningless unless the factor (or factors) on which the variable is loading is interpretable. A communality of 0.25 seems low but may be meaningful if the item is contributing to a well-defined factor. Therefore, what is important is the extent to which the item plays a role in the interpretation of the factor, though often this role is greater when communality is high (North Carolina State University, n.d.).

The broad range and lack of stability in the Control group scores for some of the HRQL domains may be accounted for by the inclusion within the relevant factors of items with low communalities. For example, the items 'thirsty' and 'tireless' have the lowest communalities of all of the items, making them candidates for exclusion from the instrument. While those items had weak statistical associations with factors 7 ('tireless' appears in this factor which is concerned with appetite), 10 ('thirsty' appears in this factor which is concerned with attention-seeking behaviour) and 12 ('thirsty' again appears in this factor which is concerned with stoicism), there was no interpretable association with the other items in those factors. This lack of a meaningful association may mean that these items introduce a high level of 'noise' into the scores for those HRQL domains, which may account for the relatively broad range of 'normal' scores for domains 10 and 12 compared with other domains, and for the relatively low stability over time of 'normal' scores for domains 7 and 12 compared with other domains. However, some other items

with communalities that are relatively low (<0.5) – 'stretching', 'cautious', 'even-tempered, 'obedient', 'confused' and 'accepting' – were those that loaded significantly onto only one factor, and therefore may be making an important contribution to the definition of the factors onto which they loaded.

An indication of the usefulness of an item to the measurement of interest is also provided by item loadings. An item's factor loading can be interpreted as the extent to which that item helps to describe that factor. An examination of the item loadings for each factor will give an indication of those items that it would be important to retain and those that may be candidates for exclusion from a refined version of the instrument.

However, the clinimetric approach, as opposed to the psychometric approach, advises that it may be important to select items not simply on account of their statistical significance but also on clinical judgement (Feinstein, 1987). Because of the risk of losing potentially valuable information, without additional evidence for the value or otherwise of individual items in contributing to the measurement of interest, the removal of any items from the GUVQuest at this early stage in its development may be unwise.

It may be appropriate to include as part of any refinement of the instrument the exploration of alternative methods of calculating scores and the effects these may have on the instrument's discriminative and evaluative ability, for example by including in calculations of HRQL domain scores only those items loading particularly heavily onto each of the factors. This was an approach taken by the developers of the Neck Pain and Disability Scale (Wheeler *et al.*, 1999), from which scores were obtained by summing only the items with loadings onto each of the multi-item factors of >0.5.

Even at this early stage in the development of the GUVQuest, however, it is appropriate to consider removing from the instrument any item that appears to be causing particular difficulty to respondents. One potential cause of difficulty with individual items is their readability. A careful examination of the ratings on individual descriptors should reveal those descriptors that some respondents may have had difficulty reading (either because these items will have been scored through by respondents, as they were instructed to do with any item about which they were 'unsure of the meaning', or because there is other evidence that the items have been misunderstood), and consideration can then be given to whether or not these items should immediately be excluded from future, refined versions of the instrument.

Chapter 6

GENERAL DISCUSSION AND CONCLUSIONS

The purpose of this research was to develop an instrument for the measurement of chronic pain in dogs. A preliminary study (Wiseman *et al.,* 2001) had provided some evidence for the anecdotal reports of the literature, that the impact of chronic pain in dogs was apparent in a wide range of behavioural domains, as it is both self-reported and observed for people. The conceptualisation of canine chronic pain as a complex, multidimensional and subjective experience suggested that that the sophisticated approaches of psychometry would be relevant to the development of an instrument to measure such pain. Using such methodology, a proxy instrument was devised to allow the dog owner to report on relevant behavioural disturbances, and evidence was obtained for that instrument's content and construct validity for the measurement of canine chronic pain by its impact upon the HRQL of the dog. The instrument was constructed of very simple items, many of which were designed to access the subjective experience of the dog, whether in chronic pain or in good health, and in that respect it offers a novel approach to the design and development of instruments for the assessment of chronic pain and HRQL in other non-verbal groups, using qualitative interpretation of behaviour by untrained observers.

Other owner-completed questionnaires have recently been proposed to measure clinical change in arthritic dogs. For example, Gingerich and Strobel (2003) included questionnaires among the range of outcome measures they used to assess treatment effects in geriatric, arthritic dogs during the clinical evaluation of a neutraceutical. In this study, the degree of disability was assessed by physical examination, by a standard questionnaire on daily activities, by a case-specific questionnaire that monitored specific impairments

affecting individual dogs, and by owner and clinician global assessments of response to therapy. The authors hypothesised that 'the attentive dog owner is capable of assessing treatment effects in conditions for which there are no consistent objective outcome markers', and sought to utilize in veterinary medicine the kinds of questionnaire-based outcome measures that have been developed in human health measurement in recent years. They based their questionnaires on validated human instruments designed to measure general and patient-specific functional impacts, but no details of the origin of their questionnaire items were provided. The results of the study revealed that, of the various measures used, the scores obtained with the patient-specific functional impairment measure and the owner's global assessment differed statistically between a treatment and a placebo group, whereas the results of the physical examination, standardised functional questionnaire and clinician global assessments did not differ statistically between groups. The authors concluded that the individualised questionnaire was sufficiently sensitive to detect treatment effects, but recommended more extensive interviews with individual owners to identify relevant functional impairments to improve the content adequacy of the instrument, and to capture owner expectations of 'normal' behaviour for their dogs.

Such owner expectations of 'normal' behaviour should be captured by the GUVQuest in the responses owners give to the items in the initial questionnaire. For a dog with a compromised health state, an owner's greater expectations of 'normal' are likely to be revealed by lower initial scores for positive descriptors, and higher initial scores for negative descriptors, than would be the case for the owner with lower expectations of 'normal' behaviour. The GUVQuest was also designed to capture relevant information from the widest possible range of relevant behavioural disturbances, recognising the potential for individual variability in the HRQL domains affected by chronic pain and the importance of content adequacy in this regard.

A more recent study than that of Gingerich and Strobel (Hielm-Björkman *et al.*, 2003) also used a range of measures to assess chronic pain, in dogs with canine hip dysplasia (CHD). These measures included a clinician-assigned locomotor index, plasma hormone assays, radiographic examination of the hip joints, and a pain assessment questionnaire for completion by the dog owner. The owners' questionnaire contained 25 questions about behaviour and locomotion, each one associated with a rating scale: most questions provided responses that were considered to be typical of a dog with chronic pain and other responses that were considered to be typical of a dog with no pain. No details were provided of the source of the questions included in this questionnaire, and so it is assumed that these were devised by the questionnaire developers, who had hypothesised that questioning owners about the changes in locomotion, behaviour and demeanour that they had observed in their dogs would provide important information regarding the identification of chronic pain. The 25-item questionnaire was completed by 41 owners of dogs with CHD and 24 owners of apparently healthy dogs. From their responses to that questionnaire, the scores on 11 questions were included in a 'chronic pain index' because they were generally applicable and provided scores that were significantly different for CHD dogs compared with control dogs. Radiographic data and physiological measurements obtained in this study could not be used reliably to indicate the presence or severity of chronic pain. Scores on the 'chronic pain index' were able to discriminate between the CHD and control dogs included in this study, although the authors recognised that the scoring system would allow other dogs to fall into a theoretical 'grey area' between the categories of having or not having chronic pain. The authors concluded that they had provided evidence for the value of the owner-completed questionnaire as part of an assessment of pain in chronic arthritis, but recognised that other behavioural variables should be considered for the development of a reliable chronic pain tool. The development of the GUVQuest began with the identification of all behavioural variables that were relevant to the measurement of chronic pain,

and the validity of this identification process was enhanced by the use of key informants.

It has been suggested that designing an instrument for use by an untrained rater in a naturally occurring situation may be considered to be a methodological strength (Breau *et al.*, 2000), and this was the intended purpose of the GUVQuest from an early stage in its development. Dog owners were considered to be key informants in the identification of relevant behavioural disturbances, and were also used to generate the collection of items that formed the item pool from which the GUVQuest's items were selected. The familiarity of these terms would improve the utility of any instrument developed for use by the community to which the lexicon belongs: the community of dog owners. There are a number of additional advantages to basing instrument items upon simple, familiar words or phrases. These include the avoidance of most of the difficulties inherent in the wording of lengthier questions or statements, as described in Chapter 3 (page 82), although readability remains a consideration. Another possible advantage of such an approach is that using rapidly understood items may facilitate access to unconscious information, a potentially valid and rich source of respondents' perceptions (Cleermans 2001; Reber and Perrig, 2001). It is interesting that in the Hospital Anxiety and Depression (HAD) scale (Zigmond and Snaith, 1983), developed to assess mood states in physically sick populations, respondents are required to select from a choice of descriptors the statement that most closely approximates to how the respondent has been feeling during the preceding week, and are instructed not to take too long over replies since 'your immediate reaction to each item will probably be more accurate than a long thought-out response'.

Because the GUVQuest's simple, largely single word items are quick to read and understand, a large number of these items was able to be included in the questionnaire, offering comprehensiveness combined with speed of completion. Because human HRQL instruments tend to be long, patients

with reduced performance status can find them difficult to complete (Caraceni *et al.*, 2002). Even without such difficulties, distilling the measurement of HRQL into a few key questions that may be rapidly answered is a goal for most instrument developers, for whom utility is a prime consideration. A common approach to achieving this goal is to develop a long instrument and then use the results of field-testing to select key questions to be included in a shorter form. The short form must then be tested by its correlation, in terms of validity and responsiveness, with the longer form. The GUVQuest samples relevant behavioural domains in a comprehensive manner by including a large number of items across a wide range of such domains. However, because the GUVQuest can be completed reasonably speedily (within 30 minutes) by most respondents, there may be no necessity to shorten the instrument and risk the loss of validity that can be associated with such a procedure. However, it is recognised that the population of owners with which the GUVQuest was pre-tested – a population of owners that had chosen to use a referral hospital – may be considered to be more highly motivated than other dog owners might be to spend time completing such a questionnaire. It would be important, for an instrument designed to be used in a veterinary practice setting, that pre-testing and field-testing be carried out in such a setting to ensure that the instrument has utility for that population, who may differ from the population attending UGSAH.

The intention of this study was to develop a generic instrument for the measurement of chronic pain in the dog, that is to say that it was intended to measure chronic pain caused by a range of chronic and painful conditions. Despite this intention, the identification of relevant behavioural domains came largely from a sample of owners of dogs diagnosed with various conditions that were classified as DJD, rather than from the wider population of dogs suffering chronic pain. Although instrument items were generated from a much wider population of dog owners, and validation of behavioural domains and descriptors was for chronic pain of any cause, to date, analysis of results of field-testing has involved only data obtained for dogs with DJD. It

is possible that some of the items selected for inclusion in the GUVQuest will be irrelevant for some other causes of chronic pain. The extent to which the GUVQuest is generic for chronic pain of any cause will be revealed by field-testing the instrument with populations suffering chronic pain of a range of causes. With an increasing population of geriatric dogs, cancer is becoming a common diagnosis in small animal practice. The condition itself may be painful and debilitating, and some therapies also have the potential to impact upon the animal's QoL. Field-testing of the GUVQuest with an oncology population will provide data and scores that can be analysed to examine the construct validity of the instrument for that population.

Should the prototype GUVQuest prove to be disease-specific for DJD, a process of adaptation from disease-specific to generic may be an option. Such a process was used to create the PedsQL™ (Varni *et al.*, 1999). This instrument provides a generic measure of paediatric HRQL although its items were initially derived from a paediatric cancer population. The generic instrument was developed by administering the original items to a new pool of patients, their families and healthcare professionals and changing, adding and deleting items as a consequence, with the final items for the PedsQL™ being selected on the basis of statistical analysis of field-testing results.

There is evidence that pain of different causes is associated with differing impacts upon each of the dimensions of pain (Price *et al.*, 1987). A recent study (Arnold *et al.*, 2004) measured the QoL of populations of people with different chronic diseases and found that these appeared to have differing relative impacts on the physical, social and psychological domains of QoL, and that these domains appeared to make different relative contributions to the patient's assessment of overall QoL. This finding supports the view that disease-specific instruments are required for sensitive measurement of HRQL and also suggests that a profile of scores for each HRQL domain may yield important information that would be obscured in a global score.

The examination of HRQL domain scores over time for individual dogs revealed that clinical change in each dog was reflected in the scores for a number of HRQL domains. It is hypothesised that the HRQL domains affected by chronic pain, and the impact upon those domains, may be influenced not only by the particular symptoms of the condition in question but also by such other variables as the duration and nature of the pain, and a range of individual and environmental factors that may influence pain perception, as is the case for human pain sufferers (Price *et al.*, 1987; Morris, 2003; Newton-John, 2003). The extent to which it is also true for dogs will be revealed by continued field-testing of the instrument with dogs suffering from a range of chronic and painful conditions of different severities and durations.

The use of a core generic measure of HRQL plus disease-specific modules has been recommended as a way to assess specific HRQL outcomes while minimising subject burden (Seid *et al.*, 1999), and this should be an aim for future development of the GUVQuest. It is expected that the current prototype will prove to contain a core element but that some of the GUVQuest's items, for example those to do with stiffness and lameness, will prove to be specific for orthopaedic chronic pain and may not be relevant to the measurement of chronic pain of certain other causes (e.g. chronic otitis externa, anal furunculosis, oral tumours). An appropriate approach to refinement of the instrument would be to identify a core set of items that prove to be generic for the assessment of chronic pain of any cause, and to develop a number of disease-specific modules containing items that are specific for chronic pain of particular causes. Similarly, the instrument may prove to contain a core set of items that are generic for the measurement of QoL, and these may form the basis of an instrument for the measurement of QoL where this is thought to be compromised by circumstances that do not include chronic pain or other impacts upon physical health.

One of the supplementary questions included in the owner questionnaire asked the owner whether or not the dog was in pain, and, if so, to estimate

how long the dog had been in pain. Where owners answered 'yes' to that question, in the initial questionnaires the median duration of pain reported by owners in the Hospital (DJD) group was 1.5 months, with three-quarters of all cases reported as having been in pain for 6 months or less and one quarter of all cases having been in pain for less than 1 month. The median duration reported in initial questionnaires by such owners in the Practice group was 4 months, with one-quarter of owners reporting durations of 3 months or less. Many of the reported durations were therefore shorter than the durations of pain specified in recent canine chronic pain studies of > 3 months (Hielm-Björkman *et al.*, 2003) or ≥1 month (Muir *et al.*, 2004). However, a comparison of clinician pain scores with owners' responses to the question of whether or not their dogs were in pain suggested that owners might be under-reporting pain, perhaps because behaviour changes associated with chronic pain can be insidious in onset, or because owners are reluctant to recognise or to admit (either consciously or unconsciously) that their dogs may be suffering pain. This latter explanation illustrates the risk of biased responses, inherent in any questionnaire instrument that has face validity. An instrument that minimises the risk of biased responses, as does the GUVQuest, and that directs attention to behavioural disturbances relevant to chronic pain, should be useful where relevant and unbiased information is sought from dog owners who may have a tendency to under-report pain. An examination of the 179 questionnaires in the 'some pain' group included in the second discriminant analysis reported in Chapter 5 (page 177) revealed that in 85 questionnaires (>47%) owners had answered 'no' to the question 'do you think your dog is in any pain?' Using HRQL domains scores calculated from owner ratings of GUVQuest items, dogs were much more frequently correctly classified (according to a clinical diagnosis of DJD) as suffering from a chronic and painful condition, providing evidence for the GUVQuest's ability to minimise respondent bias.

Response sets, the tendency to provide a certain response regardless of the question, is a danger with any questionnaire. The GUVQuest was designed to

facilitate rapid responses, and there may be a greater risk of response sets in such circumstances. However, one strategy for avoiding or at least detecting response sets, which strategy is a feature of the GUVQuest, is to include items that assess the same construct using both positive and negative items, requiring opposite scoring in each case. An examination of the patterns of scoring such items can provide evidence for response sets in data that can then be excluded from analysis (Matza *et al.*, 2004).

The problems of ceiling and floor effects, where a change in one direction or another cannot be reflected by the measure, are often a problem when measures designed for a group with restricted QoL are applied to individuals with a much better or much worse QoL (Dijkers, 1999). This has been a recognized problem in some human instruments (Ware, 1995; Wolinsky *et al.*, 1998) for example the Child Health Questionnaire (Landgraf *et al.*, 1998; Raat *et al.*, 2002) which, for that reason, may not be sensitive to change in certain individuals. For each descriptor of the GUVQuest there may be floor and ceiling effects, affecting scores for dogs at the extremes of the range, whether they are healthy or are suffering chronic pain. While this difficulty is recognised in similar instruments and is probably unavoidable, the inclusion of transition questions to accompany the core items should provide evidence of clinical change even when floor and ceiling effects are apparent.

In this study FA was used to demonstrate the construct validity of the GUVQuest. In such an analysis, there is a danger that the response to any one item in an instrument may be influenced by its proximity to another, similar item, resulting in a statistical association between the items that is an artefact and not an indication of any true association. Although the GUVQuest items were not grouped, in the instrument, into either behavioural or HRQL domains, which might have made such a problem more likely, for the purpose of improving readability they were grouped in contexts. As a consequence, some item groupings within the questionnaire were those that would be expected to be associated with a common underlying factor, with the

consequent risk just outlined. It was therefore encouraging to find that in some cases the items loading onto a factor were found in such different parts of the questionnaire that the rating of one was unlikely to have influenced the rating on the other, as might have been the case if they had been collocated. However, there may be alternative methods of presenting items that would guard against such proximity effects. For example, an approach similar to the system that is widely used in psychometric instruments for personality assessment, in which a large number of descriptors are presented in groups, and respondents are obliged to rate the descriptors within each group, could offer an alternative method of obtaining owners' ratings of the applicability of descriptors to their dogs.

It is recognised that proxy raters may be influenced in their reporting by variables such as personal expectations, stresses and mental health (Levi and Drotar, 1999; Eiser and Morse, 2001), and these may vary over time. A test-retest study to examine intra-rater reliability could be undertaken on a healthy population, but for ethical reasons it would be difficult to justify a test-retest study on a chronic pain population, since this would necessitate maintaining animals in a constantly painful condition. However, an opportunity to undertake a brief test-retest study would be offered if cases could be recruited and initial questionnaires completed a week or two before treatment was able to be commenced, for example at the time of making the hospital or practice appointment. Unfortunately, such a study could not be undertaken as part of the work reported here. Test-retest studies should be undertaken to assess the reliability of the GUVQuest, where the scores for stable subjects would be expected to be consistent over time, for example with multiple administrations of the questionnaire to owners of healthy dogs or to owners of dogs suffering chronic pain prior to treatment being commenced. Such studies can be carried out even with scales that include causal items (Fayers and Hand, 2002). In test-retest studies, the period between tests must be sufficiently long that memory effects can be ignored (Fayers and Hand, 2002). Because of the large number of items included in the GUVQuest, it is likely

to be difficult for owners to remember responses on individual items from one administration to the next, so that a week between test and retest may be sufficient. Pre-testing of such an administration could be undertaken to identify a minimum period between test and retest, before the main study is undertaken. A revised version of the GUVQuest could include some items on which ratings would be expected to remain constant, regardless of clinical condition, as an indicator of intra-observer reliability.

Responsiveness can also be assessed using a range of statistical methods, and it would be important to undertake such assessment before the GUVQuest could be considered for clinical use as an evaluative instrument. Over a series of studies on populations suffering from various conditions and using a variety of disease-specific measures with 7-point rating scales, the MID has appeared to fall consistently close to 0.5 points on a 7-point scale and it has been argued that this is a consequence of the limit of human discrimination ability (Norman *et al.*, 2003). However, others hold the view that this is too simplistic an approach (Beaton, 2003; Wright, 2003).

The interpretability of an instrument, the ease with which clinicians can identify differences in scores that correspond to trivial, small but important, moderate and large differences (degrees of improvement or deterioration), should be developed by determining how scores obtained with the instrument relate to marker states that are familiar and meaningful to clinicians. To achieve this purpose, in future field-testing of the instrument clinicians might be required to choose from various descriptions of a dog's clinical state and clinical improvement/deterioration, and could also be asked to provide free-choice descriptions of these. Analysis of these responses may help to identify those states that are familiar and meaningful to the clinician, and to relate these to HRQL domain scores and changes in these obtained with the instrument.

As well as the instrument having clinical interpretability, if it is to have value in communicating information to the owner, a different presentation of scores may be required if the owner is readily to understand the condition of the dog and changes in this, and the implications of such changes for future treatment or consideration of euthanasia.

If the validity, reliability and responsiveness of the GUVQuest, or a refined version of the instrument, are assessed as being acceptable for the clinical measurement of chronic pain, then the instrument is likely to be used by owners at home on a regular basis. This would be preferable to its completion at the time of a hospital or practice visit that may have impacts on the behaviour of the dog or the responses of the owner. Given the time-consuming nature of data capture and calculation of HRQL domain scores, it would at an early stage be important to investigate methods of automating data capture and score generation. In particular, there may be value in exploring the potential for the development of telecoms-based systems for owner response, data capture and score generation. In cooperation with individuals and companies with relevant expertise, the development of a suitable telecoms-based system could facilitate the timely input of owner ratings while in the home environment, and would speed the process of data capture and score generation for the immediate as well as longer term evaluation of the patient's HRQL. Alternatively, the use of electronic diaries, which have proved valuable in human pain assessment in the home environment (Affleck, 1996; Peters *et al.*, 2000; Aaron *et al.*, 2004) offer a tried-and-tested method of improving compliance and accuracy in such assessment, as well as providing an automated means of data capture and score generation. For example, such software can automatically check for missing data, flag irregular responses and calculate scores in real time (Wolinsky *et al.*, 1998).

If the GUVQuest is measuring chronic pain through its impact upon QoL, then the methodology may also be used to develop measures of QoL in other

circumstances in which it is at risk of being compromised, for example in long-term kennelling of rescued or quarantined dogs. In these circumstances it would be possible to carry out an examination of the inter-rater reliability of an instrument designed to be used by the dog carer, since it would be expected that in many cases there would be a number of different staff members each of whom might be expected to be familiar with the individual dog. A similar investigation of inter-rater reliability was carried out by the developers of the EDIN scale, who asked 2 nurses independently to determine EDIN scores for the same infant and who found acceptable inter-rater reliability in these circumstances (Debillon *et al.*, 2001).

An acceptance that the subjective experience of the animal is largely inaccessible to measurement has meant that animal welfare has for the past 25 years focussed on 'objective' indicators of well-being or QoL, such as longevity, reproductive success, physical integrity and various biochemical indicators. For domestic animals, informal human observation and interpretation of behaviour have long been used to assess an animal's subjective state: to judge whether or not an animal is feeling calm or aggressive, confident or fearful, is content or is suffering. The usefulness of such a 'rating' approach for the gathering of information about subtle aspects of an individual's behaviour, which is not easily obtained by other means, has now been recognised by scientists studying animal behaviour (Martin and Bateson, 1993). Recent decades of empirical research into nonverbal behaviour have revealed that humans use nonverbal cues to judge the emotional states of others, and authors have sought to explain the evolutionary origins of this important skill (Montepare, 2003; Patterson, 2003). Evidence for the validity of such human interpretations of nonverbal behaviour in other species has also been provided. In one study, observer ratings of cats' behavioural styles (e.g. aggressive, playful, sociable) were found generally to be valid and reliable (Feaver *et al.*, 1986). Qualitative judgements of the behavioural styles of individual pigs by untrained observers demonstrated significant inter-observer agreement, suggesting that such

judgements were based on commonly perceived and systematically applied criteria (Wemelsfelder *et al.*, 2000; Wemelsfelder *et al.*, 2001). Most recently, it was demonstrated that personality traits in dogs could be judged by untrained observers with 'impressive levels of accuracy' (Gosling *et al.*, 2003).

The interviews undertaken in this study provided some evidence that owners were capable of remarking and reporting styles of behaviour that they interpreted as evidence of the 'hidden' emotional or subjective states of their dogs, and degrees of and changes in such subjective states. The hypothesis that owners were capable of rating their dogs' subjective states was tested by the analysis of the data obtained with the instrument constructed upon this hypothesis. While there can be no certainly that an owner's rating of a dog's subjective state is accurate, authors including Dawkins (1980), Bateson (1991), Bekoff (1994) and Wemelsfelder (1997) have argued that it is legitimate to attempt to study such qualitative judgements scientifically.

The tension between the desire to assess animal welfare and a resistance to using qualitative methods to do so is apparent in a recent report of a survey of current practice in recognising and assessing pain, suffering and distress in UK laboratory animals (Hawkins, 2002). The author reported that currently the clinical signs used as indicators of these welfare-compromising states were largely 'subjective'. The report recommended that the message that 'subjective impressions are not necessarily correct' should be disseminated as widely as possible. It concluded that objective techniques of animal welfare measurement could have 'an immediate impact on welfare…if they are used to demonstrate that entirely subjective impressions of animal well-being are not always reliable'. However, in the same report, it was also recommended that the status of animal technicians should be high and suggested that 'everyone respects them as a valued resource – they are frequently the first to detect changes in animal behaviour including signs of suffering, so their expertise and judgement must be respected'. It also recommended ensuring that 'everyone responsible for using and monitoring animals is empathetic,

competent and confident,' since 'to interpret animal behaviour, people primarily need to be able to empathise effectively with the animals in their care'.

Increasingly, in all fields in which the measurement of QoL is important – medicine, social science, veterinary medicine and animal welfare – the central importance of the individual subject's perception of prevailing circumstances is recognised. In some of these fields, this recognition has led to advances in the measurement of pain and QoL, even for those who cannot self-report. In veterinary medicine and animal welfare, greater efforts to access the subjective experience of the animal may be overdue. Consequently, and in the absence of alternative methods, exploring the validity of qualitative judgements of the subjective experiences of others may be considered worthy of attempt (Wemelsfelder and Farish, 2004)

Those who suggest that animals do not suffer pain in the same way that people do, argue that this is because animals do not possess the cognitive capacity required for suffering – the capacity for emotion. Our results suggest that dogs with chronic pain show many of the complex behaviour changes that in people come to dominate the clinical picture over time, such as anxiety, aggression, social withdrawal and depression – changes that are, in man, considered to be associated with the emotional and evaluative dimensions of a multi-dimensional concept of pain. The evidence that chronic and painful conditions in dogs are associated with types of behaviour changes that are similar to those occurring in people suffering similar conditions, supports a similarity between dogs and man in the dimensional features of chronic pain, suggesting that the capacity for suffering exists in both species.

In 1976, an editorial in the journal *Pain* (Wall, 1976) called upon the pain research community to investigate the existence in animal species of chronic pain syndromes that are similar to those suffered by man. The present study has provided some evidence that not only the clinical signs but also the

experience of chronic and painful conditions such as osteoarthritis may be very similar for a dog and for a person. To date, however, the human pain research community has paid little attention to the resource offered by the very many (and increasing numbers of) dogs annually presenting to the veterinary community with the kinds of chronic and painful conditions that are also suffered by man (Hansen, 2003). However, in a recent study (Karai *et al.*, 2004), a canine model was used to assess the efficacy of a new surgical treatment protocol for human chronic pain control, using dogs with a clinical diagnosis of advanced cancer or osteoarthritis, for which pain control medication had proved inadequate. Assessment of improvement following treatment was made on the basis of reduced limb guarding, increased activity and improved demeanour, although no indication was given that these measures had been validated for pain assessment prior to the study in which they were used. In this case, an instrument with proven validity, reliability and responsiveness for the measurement of chronic pain would have been a valuable tool with which to assess treatment effectiveness.

The IMMPACT (Initiative on Methods, Measurement and Pain Assessment in Clinical Trials) group has recently published its recommendations on core outcome measures for chronic pain clinical trials, which include measures to assess pain, physical functioning, emotional functioning and patient's global assessment of improvement. The group did not consider for recommendation any measures for which information on the appropriateness of content, reliability, validity, responsiveness and participant burden had not been published. They acknowledged the 'important limitations of existing measures and the pressing need to develop improved methods for assessing chronic pain outcomes', and the inappropriateness of many measures in clinical trials that include cognitively impaired individuals or infants (Dworkin *et al.*, 2005).

Quality of life and pain are recognised as being subjective experiences – the most important element of each being how the individual 'feels' in the circumstances – and it has been recommended that the goal of their

measurement must be 'to gain access to the subjectivity of the participant' (Stenner *et al.*, 2003). The GUVQuest has attempted to gain such access through its use of brief, familiar terms to describe relevant behaviours and the interpretation of these as expressions of mental state by someone who is familiar with the sufferer. Even in verbal human patients, there may be some value in attempting more directly to access by self-report the subjective experience of the sufferer by focussing on simple subjective-expressive terms as an alternative to more complex and lengthy items, or items that are more concerned with the impact of the subjective experience upon functioning.

Important features of the GUVQuest are that it addresses the demonstration of good health as well as chronic pain by including both positive and negative descriptors as items. The 7-point numerical rating scale for each item represents a direct estimation approach that provides a continuum of answer options for ease of completion, maximum information, reduced error and increased efficiency. Its simple, repetitive design and its basis in an appropriate community lexicon makes it easy and quick to use by untrained raters, which are important qualities in an instrument intended to be used by individuals whose chief criterion for selection is their familiarity with the suffering individual. It facilitates the comprehensive assessment of a broad range of relevant domains, essential when measuring a complex, multi-dimensional construct such as chronic pain or HRQL, while being acceptably quick to complete. The inclusion of large numbers of positive and negative items, each associated with a ubiquitous rating scale (with inference reversed according to whether the descriptor is positive or negative), was designed to provide sensitive direct ratings of all relevant behaviours associated with the broadest possible range of health states, while minimising respondent bias.

With further evidence for the GUVQuest's validity, and a thorough assessment of its reliability and responsiveness, the design of the instrument and the process of its development may be applied to the measurement of chronic pain and HRQL in companion animals other than the dog. It may

also be appropriate to explore the potential of this approach for the measurement of chronic pain and HRQL in non-verbal human populations, such as infants and the cognitively impaired, whose inability to 'understand' their pain and the prospects of its being relieved, and whose inability to verbally communicate their suffering, present the same kinds of vulnerability and the same difficulties of pain and QoL assessment as do those species whose care is the responsibility of the veterinary profession.

REFERENCES

Aaron L.A., Mancl L., Turner J.A., Sawchuk C.N. and Klein K.M. (2004) Reasons for missing interviews in the daily electronic assessment of pain, mood and stress. *Pain* 109: 389-398.

Abu-Saad H.H. (2001) Commentary. *Archives of Disease in Childhood Fetal and Neonatal Edition* 85: F40-F41.

ACITS, The University of Texas at Austin (1995-97) Factor analysis using SAS PROC FACTOR [Online] Available at http://www.utexas.edu/cc/docs/stat53.html (last accessed 20 December 2004).

Ackerley E. (no date) Analysing and interpreting data 2 [Online] Available at http://www.maths.lancs.ac.uk/department/study/years/other/PSYCH401/lectures/wk17lect (last accessed 19 March 2005)

Affleck G., Urrows S., Tennen H., Higgins P. and Abeles M. (1996) Sequential daily relations of sleep, pain intensity, and attention to pain among women with fibromyalgia. *Pain* 68: 363-368.

Ahmedzai S. (1995) Recent clinical trials of pain control: impact on quality of life. *European Journal of Cancer* 31A(Suppl. 6): S2-S7.

American College of Veterinary Anesthesiologists (1998) American College of Veterinary Anesthesiologists position paper on the treatment of pain in animals. *Journal of the American Veterinary Medical Association* 213(5): 628-630.

American College of Veterinary Anesthesiologists (1995-2000) Position paper on the treatment of pain in animals [Online] Available at

http://www.acva.org/professional/Position/pstn.asp (last accessed 7 January 2005)

American Veterinary Medical Association (2004) *Animal Welfare Position Statements (Current as of April 2004)* [Online] Available at http://www.avma.org/policies/animalwelfare.asp (last accessed 29 June 2004)

Anand K.J.S. and Craig K.D. (1996a) New perspectives on the definition of pain. *Pain* 67: 3-6.

Anand K.J.S. and Craig K.D. (1996b) Re. Reply to letters to the editor from Merskey and Wall. *Pain* 67(1): 210.

Anil S.A., Anil L. and Deen J. (2002) Challenges of pain assessment in domestic animals. *Journal of the American Veterinary Medical Association* 220(3): 313-319.

Annett J. (2002) Subjective rating scales: science or art? *Ergonomics* 45(14): 966-987.

Argyle M. (1987) Innate and cultural aspects of human non-verbal communication. In *Mindwaves: Thoughts on Intelligence, Identity and Consciousness*, eds. Blakemore C. and Greenfield S., pp. 55-72. Oxford: Blackwell.

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.

Arnold R., Ranchor A.V., Sanderman R., Kempen G.I.J.M., Ormel J. and Suurmeijer T.P.B.M. (2004) The relative contribution of domains of quality of life to overall quality of life for different chronic diseases. *Quality of Life Research* 13: 883-896.

Asmundson G.J.G., Norton P.J. and Norton G.R. (1999) Beyond pain: the role of fear and avoidance in chronicity. *Clinical Psychology Review* 19(1): 97-119.

Association of Veterinary Teachers and Research Workers (1986) Guidelines for the recognition and assessment of pain in animals. *The Veterinary Record* 118: 334-338.

Bateson P. (1991) Assessment of pain in animals. *Animal Behaviour* 42: 827-839.

Beaton D.E., Bombardier, C., Katz J.N. and Wright J.G. (2001) A taxonomy for responsiveness. *Journal of Clinical Epidemiology* 54: 1204-1217.

Beaton D.E. (2003) Simple as possible? Or too simple? Possible limits to the universality of the one half standard deviation. *Medical Care* 41(5): 593-596.

Bech P. (1999) Health-related quality of life measurements in the assessment of pain clinic results. *Acta Anaesthesiologica Scandinavica* 43: 893-896.

Becker N., Thomsen A.B., Olsen A.K., Sjørgen P., Bech P. and Eriksen J. (1997) Pain epidemiology and health related quality of life in chronic non-malignant pain patients referred to a Danish multidisciplinary pain center. *Pain* 73: 393-400.

Beers M.H. and Berkow R. (1999-2004) *The Merck Manual of Diagnosis and therapy, Seventeenth edition, Centennial edition* [Online] Avaliable at http://www.merck.com/mrkshared/mmanual/home/jsp (last accessed 14 June 2004).

Bekoff, M. (1994) Cognitive ethology and the treatment of non-human animals: how matters of mind inform matters of welfare. *Animal Welfare* 3: 75-96.

Bekoff M. (2002) Empathy: common sense, science sense, wolves and well-being. *Behavioral and Brain Sciences* 25(1): 26-27.

Birnbacher D. (1999) Quality of life – evaluation or description? *Ethical Theory and Moral Practice* 2: 25-36.

Bowling A. (1991) *Measuring Health: a review of quality of life measurement.* Buckingham, UK: Open University Press.

Brambell F.W.R. (1965) *Report of the Technical Committee to Enquire into the Welfare of Animals Kept Under Intensive Livestock Husbandry Systems.* London, HMSO.

Brearly J.C. and Brearly M.J. (2000) Chronic pain in animals. In *Pain Management in Animals*, eds. Flecknell P.A. and Waterman-Pearson A., pp. 147-160. London: WB Saunders.

Breau L.M., McGrath P.J., Camfield C., Rosmus C. and Finley G.A. (2000) Preliminary validation of an observational pain checklist for persons with cognitive impairments and inability to communicate verbally. *Developmental Medicine and Child Neurology* 42: 609-616.

Breau L., Finlay, G.A., McGrath P.J. and Camfield C. (2002a) Validation of the Non-communicating Children's Pain Checklist–postoperative version. *Anesthesiology* 96(3): 528-535.

Breau L.M., McGrath P.J. Camfield C.S. and Finley G.A. (2002b) Psychometric properties of the non-communicating children's checklist–revised. *Pain* 99: 349-357.

Briggs A., Scott E. and Steele K. (1999) Impact of osteoarthritis and analgesic treatment on quality of life in an elderly population. *Annals of Pharmacotherapy* 33: 1154-1159.

Bronfort G. and Bouter L.M. (1999) Responsiveness of general health status in chronic low back pain: a comparison of the COOP Charts and the SF-36. *Pain* 83: 201-209.

Broom D.M. (1991) Animal welfare: concepts and measurement. *Journal of Animal Science* 69: 4167-4175.

Broström E., Hagelberg S. and Haglund-Åkerlind Y. (2004) Effect of joint injections in children with juvenile idiopathic arthritis: evaluation by 3D-gait analysis. *Acta Paediatrica* 93(7): 906-910.

Buchholz M, Karl H.W., Pomietto M. and Lynn A. (1998) Pain scores in infants: a modified infant pain scale versus visual analogue. *Journal of Pain and Symptom Management* 15: 117-124.

Burgess C. (2001) PSY2005: Factor Analysis [Online] Available at http://66.102.9.104/u/exeteruniversity?q=cache:810fAUj31VEJ:www.ex.ac.uk/Psychology/docs/courses/2016/fa/fa_lecture_handout.doc+PSY2005:+Factor+Analysis&hl=en&ie=UTF-8 (last accessed 7 January 2005).

Burghardt G.M. (1985) *Foundations of comparative ethology*. Van Nostrand Reinhold: New York.

Burghardt G.M. (1991) Cognitive ethology and critical anthropomorphism: a snake with two heads and hognose snakes that play dead. In *Cognitive ethology: the minds of other animals*, ed. Ristau, C.A., pp: 53-90. San Francisco: Erlbaum.

Burgos-Vargas R. (1999) Assessment of quality of life in children with rheumatic diseases (Editorial). *The Journal of Rheumatology* 26(7): 1432-1434.

Calaminus G. and Kiebert G. (1999) Studies on health-related quality of life in childhood cancer in the European setting: an overview. *International Journal of Cancer* 83(S12): 83-86.

Cambridge A.J., Tobias K.M., Newberry R.C. and Sarkar D.K. (2000) Subjective and objective measurements of postoperative pain in cats. *Journal of the American Veterinary Medical Association* 217(5): 685-690.

Camilleri-Brennan J. and Steele R.J.C. (1999) Measurement of quality of life in surgery. *Journal of the Royal Colleges of Surgeons of Edinburgh and Ireland* 44: 252-259.

Canadian Forest Service (2002) Principal Components and Factor Analysis [Online] Available at www.pfc.forestry.ca/profiles/wulder/mvstats/pca_fa_e.html (last accessed 29 March 2005).

Caraceni A., Cherny N., Fainsinger R., Kaasa S., Poulain P., Radbruch L. and De Conno F. (2002) Pain measurement tools and methods in clinical research in palliative care : recommendations of an expert working group of the European Association of Palliative Care. *Journal of Pain and Symptom Management* 23(3): 239-255.

Carr A.J. (1999) Beyond disability: measuring the social and personal consequences of osteoarthritis. *Osteoarthritis and Cartilage* 7: 230-238.

Chambers, C.T., Reid, G.J., McGrath, P.J. and Finley, G.A. (1996) Development and preliminary validation of a postoperative pain measure for parents. *Pain* 68: 307-313.

Chambers C.T., Finley G.A., McGrath P.J. and Walsh T.M. (2003) The parents' postoperative pain measure: replication and extension to 2-6-year-old children. *Pain* 105: 437-443.

Chapman C.R. (1976) Measurement of pain: problems and issues. In *Advances in Pain Research and Therapy Volume 1*, eds. Bonica J.J. and Albe-Fessard D., pp. 345-353. New York: Raven Press.

Chapman C.R., Casey, K.L., Dubner R., Foley K.M., Gracely R.H. and Reading A.E. (1985) Pain measurement: an overview. *Pain* 22: 1-31.

Chapman C.R. (1989) The concept of measurement: coexisting theoretical perspectives. In *Advances in Pain research and Therapy Volume 12*, eds. Chapman C.R. and Loeser J.D., p. 7. New York: Raven Press.

Chapman C.R. (1992) Suffering in animals: towards a comprehensive definition and measurement for animal care, In *Proceedings of a Conference on Animal Pain: Ethical and Scientific Perspectives*, eds. Kuchel T.R, Rose M. and Burrell J., pp. 19-25. Australian Council for the Care of Animals in Research and Testing (ANZCCART)

Christ E. (1998) *Images of Animals: Anthropomorphism and Animal Mind,* pp. 202-222. Philadelphia: Temple University Press.

Ciccetti D.V., Showalter D., Tyrer P.J. (1985) The effect of number of rating scale categories on levels of interrater reliability: a Monte Carlo investigation. *Applied Psychological Measurement* 9: 31-36.

Clark H.H. (1998) Communal lexicons. In *Context in Language Learning and Language Understanding*, eds. Malmkjær K and Williams J., pp. 60-87. Cambridge: Cambridge University Press.

Clark J.D., Rager D.R. and Calpin J.P. (1997) Animal well-being II. Stress and distress. *Laboratory Animal Science* 47(6): 571-579

Clark S.R.L. (1990) The reality of shared emotions. In *Interpretation and Explanation in the Study of Animal Behaviour Volume 1,* eds. Bekoff M. and Jamieson D., pp. 449-472. Boulder: Westview Press.

Clark, W.C., Fletcher, J.D., Janal, M.N. and Carrol, J.D. (1995) Hierarchical clustering of pain and emotion descriptors: towards a revision of the McGill Pain Questionnaire. In *Pain and the Brain: from Nociception to Cognition*, eds.

Bromm, B. and Desmedt, J. E., *Advances in Pain Research and Therapy Volume 22,* pp. 319-330. New York: Raven Press.

Clark W.C., Yang J.C., Tsui S.-L., Ng K.-F. and Clark S.B. (2002) Unidimensional pain rating scales: a multidimensional affect and pain survey (MAPS) analysis of what they really measure. *Pain* 98(3): 241-247.

Cleeland C.S. (1989) Measurement of pain by subjective report. In *Advances in Pain Research and Therapy: Issues in Pain Measurement*, eds. Chapman C.R. and Loeser J.D., pp. 391-403. New York: Raven Press.

Cleermans A. (2001) Conscious and unconscious processes in cognition. In *International Encyclopedia of the Social and Behavioral Sciences (electronic resource),* eds. Smelser N.J. and Baltes P.B., pp. 2584-2589. Amsterdam: Elsevier

Cohen R.F. (1999) Individual quality of life and assessment by carers or 'proxy respondents'. In *Individual Quality of Life: Approaches to Conceptualisation and Assessment*, eds. Joyce C.R.B., McGee H. and O'Boyle C.A, pp. 185-194. Amsterdam: Harwood Academic Publishers.

Conzemius M.G., Hill C.M., Sammarco J.L. and Perkowski S.Z. (1997) Correlation between subjective and objective measures used to determine severity of postoperative pain in dogs. *Journal of the American Veterinary Medical Association* 201(11): 1619-1622.

Cook K.F., Monahan P.O. and McHorney C.A. (2003) Delicate balance between theory and practice: health status and item response theory (editorial). *Medical Care* 41(5): 571-574.

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.

Coste J., Bouée S., Ecosse E., Leplège and Pouchot J. (2005) Methodological issues in determining the dimensionality of composite health measures using principal component analysis: case illustration and suggestions for practice. *Quality of Life Research.* In press.

Councill W., Stirling B., Bentham P., Buckley A. and Laight A. (2001) EQ-5D in patients with dementia: an investigation of inter-rater agreement. *Medical Care* 39(8): 760-771.

Craig K.D. (1999) Emotions and psychobiology. In *Textbook of Pain (4th Edition),* eds. Wall P.D. and Melzack R., pp. 331-343. London: Churchill Livingstone.

Craig K.D. and Badali M.A. (1999) On knowing an infant's pain (commentary). *Pain Forum* 8(2): 74-77.

Crombez G., Eccleston C., Baeyens F., Van Houdenhove B. And Van den Broek A. (1999) Attention to chronic pain is dependent upon pain-related fear. *Journal of Psychosomatic Research* 47(5): 403-410.

Cunningham N. (1999) Primary requirements for an ethical definition of pain (Focus). *Pain Forum* 8(2): 93-99.

Darlington R.B. (no date) *Factor Analysis* [Online] Available at http://comp9.psych.cornell.edu/Darlington/factor.htm (last accessed 20 December 2004).

Darwin C. (1872) The Expression of the Emotions in Man and Animals. London: John Murray.

Dawkins M.S. (1980) *Animal Suffering*, pp. 10-82. New York: Chapman and Hall.

Dawkins M.S. (2004) Using behaviour to assess animal welfare. *Animal Welfare* 13: S3-S7

Debillon T., Sgaggero B., Zupan V., Tres F., Magny J.-F., Bouguin M.A. and Dehan M. (1994) Sémiologie de la douleur chez le prématuré. *Archives de Pédiatrie* 1: 1085-1092.

Debillon T., Zupan V., Ravault N, Magny J.-F. and Dehan M. (2001) Development and initial validation of the EDIN scale, a new tool for assessing prolonged pain in preterm infants. *Archives of Disease in Childhood Fetal and Neonatal Edition* 85: F36-F40.

De Haes J., Curran D., Young T., Bottomly A., Flechtner H., Aaronson N., Blazeby J., Bjordal K., Brandberg Y., Greimel E., Maher J., Sprangers M. and Cull A., for the EORTC Quality of Life Study Group (2000) Quality of life evaluation in oncological clinical trials – the EORTC model. *European Journal of Cancer* 36: 821-825.

Dempster M. and Donnelly M. (2000) Measuring the health related quality of life of people with ischaemic heart disease. *Heart* 83: 641-644.

Devor M. and Seltzer Z. (1999) Pathophysiology of damaged nerves in relation to chronic pain. In *Textbook of Pain (4th Edition),* eds. Wall P.D. and Melzack R., pp 129-164. London: Churchill Livingstone.

Deyo K.S., Prkachin K.M. and Mercer S.R. (2004) Development of sensitivity to facial expression of pain. *Pain* 107: 16-21.

Dijkers M. (1999) Measuring quality of life: methodological issues. *American Journal of Physical Medicine and Rehabilitation* 78(3): 286-300.

Duncan I.J.H. and Fraser D. (1997) Understanding Animal Welfare. In *Animal Welfare*, eds. Appleby M.C. and Hughes B.O., pp. 19-31. Wallingford: CAB International.

Dworkin R.H., Turk D.C., Farrar J.T., Haythornthwaite J.A., Jensen M.P., Katz N.P., Kerns R.D., Stucki G., Allen R.R., Bellamy N., Carr D.B., Chandler J., Cowan P., Dionne R., Galer B.S., Hertz S., Jadad A.R., Kramer L.D., Manning D.C., Quessy S., Rappaport B.A., Robbins W., Robinson J.P., Rothman M., Royal M.A., Simon L., Stauffer J.W., Stein W., Tollett J., Wernicke J. and Witter J. (2005) Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain* 113: 9-19.

Eibl-Eibesfeldt I (1972) Similarities and differences between cultures in expresssive movements. In *Non-verbal Communication*, ed. Hinde R., pp. 297-312. Cambridge: Cambridge University Press.

Eisenberger, N., Lieberman, M.D. and Williams, K.D. (2003) Does rejection hurt? An fMRI study of social exclusion. *Science* 302(5643): 290-292.

Eiser C. and Morse R. (2001) A review of measures of quality of life for children with chronic illness. *Archives of Disease in Childhood* 84: 205-211.

European Federation of IASP Chapters (2004). EFIC's declaration on pain as a health problem, a disease in its own right [Online] Available at http://www.efic.org/about_pain.htm#efic_declaration (last accessed 25 January 2005).

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

Fayers P.M., Hopwood P., Harvey A., Girling D.J., Machin D. and Stephens R., on behalf of the MRC Cancer Trials Office (1997) Quality of life assessment in clinical trials – guidelines and a checklist for protocol writers: the U.K. Medical Research Council experience. *European Journal of Cancer* 33(1): 20-28.

Fayers P.M and Hand D.J. (2002) Causal variables, indicator variables and measurements scales: an example from quality of life. *Journal of the Royal Statistical Society* 165(2): 1-21.

Feaver J., Mendl M. and Bateson P. (1986) A method for rating the individual distinctiveness of domestic cats. *Animal Behaviour* 34: 1016-1025.

Feeny D., Furlong W., Mulhern R.K., Barr R.D. and Hudson M. (1999) A framework for assessing health-related quality of life among children with cancer. *International Journal of Cancer* 83(S12): 2-9.

Feinstein A.R. (1987) *Clinimetrics*, pp. 249-253. Yale: Yale University.

Fields H.L. and Basbaum A.I. (1999) Central nervous system mechanisms of pain modulation. In *Textbook of Pain (4th Edition),* eds. Wall P.D. and Melzack R., pp 309-329. London: Churchill Livingstone.

Firth A.M. and Haldane S.L. (1999) Development of a scale to evaluate postoperative pain in dogs. *Journal of the American Veterinary Medical Association* 214(5): 651-659.

Flecknell, P.A. (1985) Recognition and alleviation of pain in animals. In *Advances in Animal Welfare Science 1985,* eds. Fox M.W. and Mickley L.D., pp. 61-77. Dordrecht: Martinus Nijhoff Publishers.

Flecknell P.A. and Molony V. (1997) Pain and Injury. In *Animal Welfare*, eds. Appleby M.C. and Hughes B.O., pp. 64-69 Oxford: CAB International.

Franck L.S., Greenberg C.S. and Stevens B. (2000) Pain assessment in infants and children. *Pediatric Clinics of North America* 47(3): 487-512.

Fraser D. and Matthews L.R. (1997) Preference and motivation testing. In *Animal Welfare*, eds. Appleby M.C. and Hughes B.O., pp. 159-173. Wallingford: CAB International.

Fraser D., Weary D.M., Pajor E.A., Milligan, B.N. (1997) A scientific conception of animal welfare that reflects ethical concerns. *Animal Welfare* 6: 187-205.

Friendly M. (1995) Planning a Factor Analytic Study [Online] Available at http://www.psych.yorku.ca/lab/psy6140/fa/facplan.htm (last accessed 20 December 2004).

Fujita F., Lahmann B., Otsuka K., Lyass S., Hiatt J.R. and Phillips E.H. (2004) Quantification of pain and satisfaction following laparoscopic and open hernia repair. *Archives of Surgery* 139: 596-600.

Garratt A., Schmidt L., Mackintosh A. and Fitzpatrick R. (2002) Quality of life measurement: bibliographic study of patient assessed health outcome measures. *British Medical Journal* 324: 1-5.

Garrod Prof. S., Dept of Psychology, University of Glasgow, UK.

Gatchel R.J., Mayer T., Dersch J., Robinson R. and Polatin P. (1999) The association of the SF-36 Health Status Survey with 1-year socio-economic outcomes in a chronically disabled spinal disorder population. *Spine* 24(20): 2162-2170.

Gauvain-Piquard, A., Rodary, C., Rezvani, A. and Lemerle, J. (1987) Pain in children aged 2–6 years: a new observational rating scale elaborated in a pediatric oncology unit – preliminary report. *Pain* 31: 177-188.

Gauvain-Piquard A., Rodary C., Rezvani A. and Serbouti S. (1999) The development of the DEGR®: a scale to assess pain in young children with cancer. *European Journal of Pain* 3: 165-176.

Geisser M.E., Roth R.S., Theisen M.E., Robinson M.E. and Riley J.L. (2000) Negative affect, self report of depressive symptoms, and clinical depression:

relation to the experience of chronic pain. *The Clinical Journal of Pain* 16: 110-120.

Gibson D.R., Hudes E.S. and Donovan D. (1999) Estimating and correcting for response bias in self-reported HIV risk behavior. *Journal of Sex Research* 36(1): 96-101.

Gilchrist V.J. (1999) Key Informant Interviews. In *Qualitative Research Volume 1*, eds. Bryman A. and Burgess R.G., pp. 354-371. London: Sage Publications.

Gill T.M. (1995) Quality of life assessment: values and pitfalls. *Journal of the Royal Society of Medicine* 88: 680-682.

Gingerich D.A. and Strobel J.D. (2003) Use of client-specific outcome measures to assess treatment effects in geriatric, arthritic dogs: controlled clinical evaluation of a neutraceutical. *Veterinary Therapeutics* 4(4): 376-386.

Gosling S.D., Kwan V.S.Y. and John O.P. (2003) A dog's got personality: a cross-species comparative approach to personality judgments in dogs and humans. *Journal of Personality and Social Psychology* 85(6): 1161-1169.

Graue M., Wentzel-Larsen T., Hanestad B.R., Båtsvik B. and Søvik O. (2003) Measuring self-reported, health-related, quality of life in adolescents with type 1 diabetes using both generic and disease-specific instruments. *Acta Pædiatrica* 92: 1190-1196.

Griffin D.R. (1992) *Animal Minds,* pp. 26-27. Chicago: The University of Chicago.

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

Guyatt G.H., Feeny D.H. and Patrick D.L. (1993) Measuring health-related quality of life. *Annals of Internal Medicine* 118: 622-629.

Guyatt G.H. Osoba D., Wu A., Wyrwich K.W. and Norman G.R.; the Clinical Significance Consensus Meeting Group (2002) Methods to explain the clinical significance of health status measures. *Mayo Clinic Proceedings* 77(4): 371-383.

Hadjistavropousos T. and Craig K.D. (2002) A theoretical framework for understanding self-report and observational measures of pain: a communications model. *Behaviour Research and Therapy* 40: 551-570.

Hansen B.D. (2003) Assessment of pain in dogs: veterinary clinical studies. *ILAR Journal* 44(3): 197-205.

Hardie EM, Hansen BD and Carroll GS (1997) Behavior after ovariohysterectomy in the dog: what's normal? *Applied Animal Behaviour Therapy* 51: 111-128.

Hare B., Brown M., Williamson C. and Tomasello M. (2002) The domestication of social cognition in dogs. *Science* 298(5598): 1634-1636.

Harstall C. and Ospina M. (2003) How prevalent is chronic pain? *Pain Clinical Updates (International Association for the Study of Pain)* XI(2): 1-4.

Hawkins P. (2002) Recognizing and assessing pain, suffering and distress in laboratory animals: a survey of current practice in the UK with recommendations. *Laboratory Animals* 36: 378-395.

Hellyer P.W. (2002) Treatment of pain in dogs and cats. *Journal of the American Veterinary Medical Association* 221(2): 212-215.

Hielm-Björkman A.K., Kuusela E., Lipman A., Markkola A., Saarto E., Huttunen P., Leppäluoto J., Tulamo R.-M. and Raekallio M. (2003)

Evaluation of methods for assessment of pain associated with chronic osteoarthritis in dogs. *Journal of the American Veterinary Medical Association* 222(11): 1552-1558.

Holton L., Scott E.M., Nolan A.M., Reid J., Welsh E and Pawson P. (1997) The development of a composite measure scale to assess pain in dogs. *Journal of Veterinary Pharmacology & Theraputics* 20 (Suppl.1): 167-168.

Holton L.L., Scott E.M., Nolan A.M., Reid J., Welsh E. and Flaherty D. (1998) Comparison of three methods used for assessment of pain in dogs. *Journal of the American Veterinary Medical Association* 212: 61–66

Holton L., Reid J., Scott E.M., Pawson P. and Nolan A. (2001) Development of a behaviour-based scale to measure acute pain in dogs. *Veterinary Record* 148: 525–531.

Hsu Y. and Serpell J.A. (2003) Development and validation of a questionnaire for measuring behavior and temperament traits in pet dogs. *Journal of the American Veterinary Medical Association* 223(9): 1293-1300.

Hsu Dr Y., Department of Biology, National Taiwan Normal University, Taipei, Taiwan.

Huntingford F. (1984) *The Study of Animal Behaviour.* London: Chapman & Hall.

Hutchison, A., MRCVS, Dalblair Veterinary Surgery, Ayr, UK.

IASP Subcommittee on Taxonomy (1979) Pain terms: a list with definitions and notes on usage. *Pain* 6: 247-252.

IASP Subcommittee on Taxonomy (1986) Classification of chronic pain: descriptions of chronic pain syndromes and definitions of pain terms. *Pain* Supp 3: S1-S225.

Insights Learning and Development Ltd. (1992-2003) [Online] Available at http://www.insights.com (Last accessed 30 January 2005)

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

Johnston C.C. (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 G.A. and McGrath P.J., pp 5-20. Seattle: IASP Press.

Johnston V.S. (1999) *Why we feel: the science of human emotions.* Reading, Mass: Perseus books, p. 87, quoted in McMillan F.D. and Rollin B.E. (2001) The presence of mind: on reunifying the animal mind and body *Journal of the American Veterinary Medical Association* 218(11): 1723-1727.

Juniper E.F., Guyatt G.H. and Jaeschke R. (1996) How to develop and validate a new health-related quality of life instrument. In *Quality of Life and Pharmacoeconomics in Clinical Trials (2nd Edition)*, ed. Spilker B., pp. 49-56. Philadelphia: Lippincott-Raven Publishers.

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.

Kappesser J. and Williams A.C.deC. (2002) Pain and negative emotions in the face: judgments by health care professionals. *Pain* 99:197-206.

Karai L., Brown D.C., Mannes A.J., Connelly S.T., Brown J., Gandai M., Wellisch O.M., Neubert J.K., Olah Z. and Ladarola M.J. (2004) Deletion of banilloid receptor 1 – expressing primary afferent neurons for pain control. *The Journal of Clinical Investigation* 113(9): 1344-1352.

Katz N. (2002) The impact of pain management on quality of life. *Journal of Pain and Symptom Management* 24(1S): S38-S47.

Keefe F.J., Lumley M. Anderson T., Lynch T. and Carson K.L. (2001) Pain and emotions: new research directions. *Journal of Clinical Psychology* 57(4): 587-607.

Keeling L. and Jensen P. (2002) Behavioural disturbances, stress and welfare. In *The Ethology of Domestic Animals*, ed. Jensen P., pp. 79-98. Wallingford: CABI Publishing.

Kitchen H., Aronson A.L., Bittle J.L., McPherson C.W., Morton D.B., Pakes S.P. Rollin B., Rowan A.N., Sechzer J.A., vanderlip J.E., Will J.A., Clark A.S. and Gloyd J.S. (1987) Panel report on the colloquium on recognition and alleviation of animal pain and distress. *Journal of the American Veterinary Medical Association* 191(10): 1186-1191.

Knotkova H., Clark W.C., Mokrejs P., Padour F. and Kuhl J. (2004) What do ratings on unidimensional pain and emotion scales really mean? A multidimensional affect and pain survey (MAPS) analysis of cancer patient responses. *Journal of Pain and Symptom Management* 28(1): 19-27.

Lamont L.A., Tranquilly W.J. and Grimm K.A. (2000) Physiology of pain. In *Veterinary Clinics of North America: Small Animal Practice Volume 30(4), Management of Pain*, ed. Mathews K.A., pp. 703-727. Philadelphia: W.B.Saunders.

Landgraf J.M. and Abetz L.N. (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.

Landgraf J.M., Maunsell E., Nixon Speechley K., Bullinger M., Campbell S., Abetz L. and Ware J.E. (1998) Canadian-French, German and UK versions of the Child Health Questionnaire: methodology and preliminary item scaling results. *Quality of Life Research* 7: 433-445.

Landgraf J.M. (1999) Measuring health-related quality of life in pediatric oncology patients: a brief commentary on the state of the art of measurement and application (discussion). *International Journal of Cancer* Supplement 12: 147-150.

Lascelles B.D.X. and Main D.C.J. (2002) Surgical trauma and chronically painful conditions – within our comfort level but beyond theirs? *Journal of the American Veterinary Medical Association* 221(2): 215-222.

Lee V.C. and Rowlingson J.C. (1996) Defining quality of life in chronic pain. In *Quality of Life and Pharmacoeconomics in Clinical Trials (2nd Edition)*, ed. Spilker B., pp. 853-864. Philadelphia: Lippincott-Raven Publishers.

Lester P. and Gaynor J.S. (2000) Management of cancer pain. In *Veterinary Clinics of North America: Small Animal Practice Volume 30(4), Management of Pain,* ed. Mathews K.A., pp. 951-966. Philadelphia: WB Saunders.

Levi R.B. and Drotar D. (1999) Health-related quality of life in childhood cancer: discrepancy in parent-child reports. *International Journal of Cancer* 12: 58-64.

Liang M.H. (2000) Longitudinal construct validity: establishment of clinical meaning in patient evaluative instruments. *Medical Care* 38(9) Supplement II: II-84–II-90.

Liang M.H., Lew R.A., Stucki G, Fortin P.R. and Daltroy L. (2002) Measuring clinically important changes with patient oriented questionnaires. *Medical Care* 40(4) Supplement II: II-45–II-51.

Lindley S. BVSc MRCVS, Small Animal Hospital, University of Glasgow, UK.

Lund V. and Rocklinsberg H. (2001) Outlining a conception of animal welfare for organic farming systems. *Journal of Agricultural and Environmental Ethics* 14(4): 391-424.

Luscombe M.D. and Williams J.L. (2003) Comparison of a long spinal board and vacuum mattress for spinal immobilisation. *Emergency Medicine* 20: 476-478.

McClain B.C. (2002) Measurement of pain in children: state-of the art considerations. *Anesthesiology* 96(3): 523-526.

McCracken G. (1988) *The Long Interview*, pp. 21-46. Thousand Oaks, Calif.: Sage Publications.

McDowell I. and Newell C. (1996) *Measuring Health: a Guide to Rating Scales and Questionnaires (2nd Edition),* pp. 16-18. Oxford: Oxford University Press.

McGrath P.J., Johnson G., Goodman J.T., Schillinger J., Dunn J. and Chapman J-A. (1985) Cheops: a behavioural scale for rating postoperative pain in children. In *Advances in Pain Research and Therapy Volume 9*, ed. Fields H.L., pp. 395-402. New York: Raven Press.

McGrath, P.J., Rosmus, C., Canfield, C., Campbell, M.A. and Hennigar, A. (1998) Behaviours caregivers use to determine pain in non-verbal, cognitively impaired individuals. *Developmental Medicine and Child Neurology* 40: 340-343.

McGrath P.J. and Unruh A.M. (1999) Measurement and assessment of paediatric pain. In *Textbook of Pain (4th Edition),* eds. Wall P.D. and Melzack R., pp. 371-384. Edinburgh: Churchill Livingstone.

McLaughlin, R (2000) Management of chronic osteoarthritic pain. *Veterinary Clinics of North America* 30(4): 933-949.

McMillan F.D. (2000) Quality of life in animals. *Journal of the American Veterinary Medical Association* 216(12): 1904-1910.

McMillan F.D. and Rollin B.E. (2001) The presence of mind: on reunifying the animal mind and body. *Journal of the American Veterinary Medical Association* 218(11): 1723-1727.

McMillan F.D. (2003a) A world of hurts – is pain special? *Journal of the American Veterinary Medical Association* 223(2): 183-186.

McMillan F.D. (2003b) Maximising quality of life in ill animals. *Journal of the American Animal Hospital Association* 39: 227-235.

McPherson C.J. and Addington-Hall J.M. (2003) Judging the quality of care at the end of life: can proxies provide reliable information? *Social Science and Medicine* 56(1): 95-109.

Mama K. (1999) New options for managing pain in small animals. *Veterinary Medicine* 94(4): 352-356.

Manificat S., Dazord A., Langue J., Danjou G., Bauche P., Bovet F., Cubells J.M., Luchelli R., Tockert E. and Conway K. (1999) A new instrument to evaluate infant quality of life. *MAPI Research institute Quality of Life Newsletter* 23: 7-8.

Martin P, Bateson P. (1993) *Measuring behaviour: an Introductory Guide (2nd Edition),* pp. 80-83. Cambridge: Cambridge University Press.

Mathews K.A. (2000) Pain assessment and general approach to management. *Veterinary Clinics of North America – Small Animal Practice* 30(4): 729-755.

Matza L.S., Swensen A.R., Flood E.M., Secnik K. and Leidy N.K. (2004) Assessment of health-related quality of life in children: a review of conceptual, methodological, and regulatory issues. *Value in Health* 7(1): 79-92.

Medical Outcomes Trust (2001) [online] Available at http://www.outcomes-trust.org/instruments.htm (last accessed 20 February 2005).

Meenan R.F., Gertman P.M., Mason J.H. and Dunaif R. (1982) The Arthritis Impact Measurement Scales: further investigations of a health status measure. *Arthritis and Rheumatism* 25(9): 1048-1053.

Meenan R.F., Mason J.H., Anderson J.A., Guccione A.A. and Kazis L.E. (1992) AIMS2: The content and properties of a revised and expanded Arthritis Impact Measurement Scales health status questionnaire. *Arthritis and Rheumatism* 35(1): 1-10.

Melzack R. (1975) The McGill Pain Questionnaire: major properties and scoring methods. *Pain* 1: 277-299.

Melzack R. and Casey K. L. (1968) Sensory, motivational and central control determinants of pain: a new conceptual model. In *The Skin Senses*, ed. Kenshalo D., pp. 423-439. Springfield, IL: Chas. C. Thomas.

Melzack R. and Katz J. (1992) The McGill Pain Questionnaire: appraisal and current status. In *Handbook of Pain Assessment,* eds. Turk D.C. and Melzack R., pp. 152-168. New York: The Guilford Press.

Melzak R. and Katz J. (1999) Pain measurement in persons in pain. In *Textbook of Pain (4th Edition)*, eds. Wall P. D. and Melzack R., pp. 409-426. Edinburgh: Churchill Livingstone.

Melzack R. and Wall P.D. (1965) Pain mechanisms: a new theory. *Science* 150 (3669): 971-979.

Melzack R. and Wall P.D. (1991) *The Challenge of Pain*, pp. 7-8. London: Penguin Books.

Mench J.A. and Mason G.J. (1997) Behaviour. In *Animal Welfare*, eds. Appleby M.C. and Hughes B.O., pp. 127-141. Wallingford: CAB International.

Merikle P.M. and Daneman M. (1998) Psychological investigations of unconscious perception. *Journal of Consciousness Studies* 5: 5-18.

Merskey H. and Bogduk N. (1994) *Classification of Chronic Pain (2nd Edition)*, pp. xi-xii. Seattle: IASP Press.

Meyer, R.A., Campbell, J.N., & Raja, S.N. (1994). Peripheral and neural mechanisms of nociception. In *Textbook of Pain (3rd Edition),* eds. Wall P.D. and Melzack R., pp. 13-44. Edinburgh: Churchill Livingstone.

Miklósi A, Topál J. and Csányi V. (2004) Comparative social cognition: what can dogs teach us? *Animal Behaviour* 67: 995-1004.

Miller G.A. (1956) The magical number seven, plus or minus two: some limits on our capacity for processing information. *The Psychological Review* 63(2): 81-97.

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

Molony V.J. and Kent J.E. (1997) Assessment of acute pain in farm animals using behavioural and physiological measurements. *Journal of Animal Science* 75: 266-272.

Montepare J.M. (2003) Introduction. Evolution and nonverbal behavior: adaptive social perceptions. *Journal of Nonverbal Behavior* 27(2): 61-64.

Moore D.S. (1991) *Statistics: Concepts and Controversies (3rd Edition)*, pp. 134-155. New York: W.H.Freeman & Co.

Morgan C.L (1894) *An Introduction to Comparative Psychology (2nd Edition)*, p. 53. New York: Scribners.

Morris D.B. (2003) The challenge of pain and suffering. In *Clinical Pain Management: Chronic Pain*, eds. Jensen T.S., Wilson P.R. and Rice A.S.C., pp. 3-13. London: Arnold.

Morton D.B. and Griffiths P.H.M. (1985) Guidelines on the recognition of pain, distress and discomfort in experimental animals and an hypothesis for assessment. *The Veterinary Record* 116: 431-436.

Muir W.W. III (2002) Physiology and pathophysiology of pain. In *Handbook of Veterinary Pain Management,* eds. Gaynor J.S. and Muir W.W. III, pp. 30-33. St Louis: Mosby.

Muir W.W. III and Gaynor J.S. (2002) Pain behaviors. In *Handbook of Veterinary Pain Management,* eds. Gaynor J.S. and Muir W.W. III, pp. 65-81. St Louis: Mosby.

Muir W.W. III, Wiese A.J. and Wittum T.E. (2004) prevalence and characteristics of pain in dogs and cats examined as outpatients at a veterinary teaching hospital. *Journal of the American Veterinary Medical Association* 224(9): 1459-1463.

Narsinghani U. and Anand K.J.S. (2000) Developmental neurobiology of pain in neonatal rats. *Lab Animal* 29(9): 29-39.

Naughton M.J., Schumaker S.A., Anderson R.T. and Czajkowski S.M. (1996) Psychological aspects of health-related quality of life measurement: tests and scales. In *Quality of Life and Pharmacoeconomics in Clinical Trials (2nd Edition)*, ed. Spilker B., pp. 117-131. Philadelphia: Lippincott-Raven Publishers.

Newton-John T.R. (2003) Psychological effects of chronic pain and their assessment in adults. In *Clinical Pain Management: Chronic Pain*, eds. Jensen T.S., Wilson P.R. and Rice A.S.C., pp. 101-111. London: Arnold.

Ng M.M.L., Leung M.C.P. and Poon D.M.Y. (2003) The effects of electro-acupuncture and transcutaneous electrical nerve stimulation on patients with painful osteoarthritic knees: a randomized controlled trial with follow-up evaluation. *The Journal of Alternative and Complementary Medicine* 9(5): 641-649.

Norman N.G., Sloan J.A. and Wyrwich K.W. (2003) Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Medical Care* 41(5): 582-592.

North Carolina State University (no date) *Factor Analysis* [Online] Available at http://www2.chass.ncsu.edu/garson/pa765/factor.htm (last accessed 16 April 2004).

Norton P.J., Asmundson G.J.G., Norton G.R. and Craig K.D. (1999) Growing pain: 10-year research trends in the study of chronic pain and headache. *Pain* 79: 59-65.

Nunally J.C. and Bernstein I.H. (1994) *Psychometric Theory*, pp. 3-30. New York: London: McGraw-Hill.

Padilla G.V., Grant M.M., Ferrell B.R. and Presant C. (1996). Quality of Life - Cancer. In *Quality of Life and Pharmacoeconomics in Clinical Trials (2nd Edition)*, ed. Spilker, B., pp. 301-308. New York: Raven Press.

Panksepp J. (1998) *Affective neuroscience: the foundations of human and animal emotions.* New York: Oxford University Press, p. 182, quoted in McMillan F.D. and Rollin B.E. (2001) The presence of mind: on reunifying the animal mind and body. *Journal of the American Veterinary Medical Association* 218(11): 1723-1727.

Patrick D.L and Chiang Y-p (2000) measurement of health outcomes in treatment effectiveness evaluations: conceptual and methodological challenges. *Medical Care* 38(9) Supplement II: II-14–II-25.

Patterson M.L. (2003) Commentary. Evolution and nonverbal behavior: functions and mediating processes. *Journal of Nonverbal Behavior* 27(3): 201-207.

Paul-Murphy J., Ludders J.W., Robertson S.A., Gaynor J.S., Hellyer P.W. and Wong P.L. (2004) The need for a cross-species approach to the study of pain in animals. *Journal of the American Medical Association* 224(5): 692-697.

Payne S.L.. (1951)*The Art of Asking Questions*, pp. 228-237. Princeton, NJ: Princeton University Press.

Penny K.I., Purves A.M., Smith B.H., Chambers W.A. and Smith W.C. (1999) Relationship between the chronic pain grade and measures of physical, social and psychological well-being. *Pain* 79: 275-279.

Peters M.L., Sorbi M.J., Kruise D.A., Kerssens J.J., Verhaak P.F.M. and bensing J.M. (2000) Electronic diary assessment of pain, disability and psychological adaptation in patients differing in duration of pain. *Pain* 84: 181-192.

Peterson N.C. (2004) Assessment of pain scoring. *Contemporary Topics* 45(1): 75-76.

Preston C.C. and Coleman A.M. (2000) Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica (Amst)* 104: 1-15.

Preston S.D. and de Waal F.B.M. (2002) Empathy: its ultimate and proximate bases. *Behavioral and Brain Sciences* 25: 1-72.

Price D.D., Harkins S.W. and Baker C. (1987) Sensory-affective relationships among different types of clinical and experimental pain. *Pain* 28: 297-307.

Prieto E.J and Geisinger K.F. (1983) Factor-analytic studies of the McGill Pain Questionnaire. In *Pain Measurement and Assessment,* ed. Melzack R., pp. 63-70. New York: Raven Press.

Prkachin K.M. and Craig K.D. (1995) Expressing pain: the communication and interpretation of facial pain signals. *Journal of Nonverbal Behaviour* 19(4): 191-205.

Prkachin K.M., Schultz I., Berkowitz J., Hughes E. and Hunt D. (2002) Assessing pain behaviour of low-back pain patients in real time: concurrent validity and examiner sensitivity. *Behaviour Research and Therapy* 40: 595–607

Raat H., Bonsel G.J., Essink-Bot M.-L., Landgraf J.M. and Gemke R.J.B.J. (2002) Reliability and validity of comprehensive health status measures in children: the Child Health Questionnaire in relation to the Health Utilities Index. *Journal of Clinical Epidemiology* 55: 67-76.

Ramelet A.-S., Abu-Saad H.H., Rees N. and McDonald S. (2004) The challenges of pain measurement in critically ill young children: a comprehensive review. *Australian Critical Care* 17(1): 33-45.

Rasmussen J.L. and Rajecki D.W. Differences and similarities in humans' perceptions of the thinking and feeling of a dog and a boy (1995) *Society and Animals* 3(2): 117-137.

Reber R. and Perrig W. (2001) Psychology of perception without awareness. In *International Encyclopedia of the Social and Behavioral Sciences (electronic resource),* eds. Smelser N.J. and Baltes P.B., pp. 11215-11218. Amsterdam: Elsevier.

Reese C.J., Trotter E.J., Short C.E., Erb H.N. and Barlow L.L. (2000) Assessing the efficacy of perioperative carprofen administration in dogs

undergoing surgical repair of a ruptured cranial cruciate ligament. *Journal of the American Animal Hospital Association* 36: 448-455.

Reid, G.J., Hebb, J.P.O., McGrath, P.J., Finley, G.A. and Forward, S.P. (1995) Cues parents use to assess postoperative pain in their children. *The Clinical Journal of Pain* 11: 229-235.

Robertson S.A. (2002) What is pain? *Journal of the American Veterinary Medical Association* 221:202–205.

Robertson S.A. (2003) How do we know if they hurt? Pain assessment in small animals. *Veterinary Medicine* 98(8): 700-708.

Rogler L.H., Mroczek D.K., Fellows M. and Loftus S.T. (2001) The neglect of response bias in mental health research. *Journal of Nervous and Mental Disease* 189(3): 182-187.

Rollin B.E. (1985) Animal Pain. In *Advances in Animal Welfare Science*, eds. Fox M.W. and Mickley L.D., pp. 91-105. Dordrecht, NL: Martinus Nijhoff Publishers.

Royal College of Veterinary Surgeons (2005) *Guide to Professional Conduct* [Online] Available at http://www.rcvs.org.uk/Templates/Internal.asp?NodeID=89642 (last accessed 7 January 2005)

Russo C.M. and Brose W.G. (1998) Chronic pain. *Annual Reviews Medicine* 49: 123-133.

Rutherford K.M.D. (2002) Assessing pain in animals. *Animal Welfare* 11: 31-53.

Sandvik E., Diener E. and Seidlitz L. (1993) Subjective well-being: the convergence and stability of self-report and non-self-report measures. *Journal of Personality* 61(3): 317-342.

Schilab T.S.S. (2002) Anthopomorphism and mental state attribution. *Animal Behaviour* 63:1021-1026.

Schipper H., Clinch J. McMurray A. and Levitt M. (1984) Measuring the quality of life of cancer patients: the Functional Living Index–Cancer: development and validation. *Journal of Clinical Oncology* 2(5): 472-483.

Schipper H. (1990) Quality of life: meaning and measurement. In *Advances in Pain Research and Therapy Volume 16*, ed. Foley K.M., pp. 89-98. New York: Raven Press.

Schlenk E.A., Erlen J.A., Dunbar-Jacob J., McDowell J., Engberg S., Sereika S.M., Rohay J.M. and Bernier M.J. (1998) Health-related quality of life in chronic disorders: a comparison across studies using the MOS SF-36. *Quality of Life Research* 7: 57-65.

Schwarz N., Knäuper B., Hippler H-J., Noelle-Neumann E. and Clark L. (1991) Rating scales: numeric values may change the meaning of scale labels. *Public Opinion Quarterly* 55: 570-582.

Seid M., Varni J.W., Rode C.A. and Katz E.R. (1999) The Pediatric Cancer Quality of Life Inventory: a modular approach to measuring health-related quality of life in children with cancer. *International Journal of Cancer* 83(S12): 71-76

Serlin R.C., Mendoza T.R., Nakamura Y., Edwards K.R. and Cleeland C.S. (1995) When is cancer pain mild, moderate of severe? Grading pain severity by its interference with function. *Pain* 61: 277-284.

Serpell J.A. and Hsu Y. (2001) Development and validation of a novel method for evaluating behavior and temperament in guide dogs. *Applied Animal Behaviour Science* 72: 347-364.

Short C.E. (1998) Fundamentals of pain perception in animals. *Applied Animal Behaviour Science* 59: 125-133.

Skevington S.M., MacArthur P. and Somerset M. (1997) Developing items for the WHOQOL: an investigation of contemporary beliefs about quality of life related to health in Britain. *British Journal of Health Psychology* 2: 55-72.

Skevington S.M. (1998) Investigating the relationship between pain and discomfort and quality of life, using the WHOQOL. *Pain* 76: 395-406.

Skevington S.M., Carse M.S., Williams A.C.de.C. (2001) Validation of the WHOQOL-100: pain management improves quality of life for chronic pain patients. *The Clinical Journal of Pain* 17: 264-275.

Sneeuw K.C.A., Aaronson N.K., Sprangers M.A.G., Detmar S.B., Wever L.D.V. and Schornagel J.H. (1999) Evaluating the quality of life of cancer patients: assessments by patients, significant others, physicians and nurses. *British Journal of Cancer* 81: 87-94.

Sneeuw K.C.A., Sprangers M.A.G. and Aaronson N.K. (2002) The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease. *Journal of Clinical Epidemiology* 55: 1130–1143.

Soma L.R. (1985) Behavioral changes and the assessment of pain in animals. In *Proceedings of the 2nd International Congress of Veterinary Anesthesia, Sacramento, CA.*, pp. 38-41.

Sprangers M.A.G. and Aaronson N.K. (1992) The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease: a review. *Journal of Clinical Epidemiology* 45(7): 743-760.

Stallard P, Williams L, Velleman R, Lenton S., McGrath P.J. and Taylor G. (2002) The development and evaluation of the pain indicator for cognitively impaired children (PICIC). *Pain* 98: 145-149.

StatSoft (1984-2003) Principal Components and Factor Analysis [Online] Available at http://www.statsoft.com/textbook/stfacan.html (last accessed 20 December 2004).

Stenner P.H.D., Cooper D. and Skevington S.M. (2003) Putting the Q into quality of life; the identification of subjective constructions of health-related quality of life using Q methodology. *Social Science and Medicine* 57: 2161-2172.

Sternbach R.A. (1981) Chronic pain as a disease entity. *Triangle* 20: 27-32.

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

Streiner D.L. and Norman G.R. (1995) *Health measurement scales. A practical guide to their development and use (2nd Edition),* pp. 1-180. New York: Oxford University Press.

Stroud M.W., Thorn B.E, Jensen M.P. and Boothby J.L. (2000) The relation between pain beliefs, negative thoughts, and psychosocial functioning in chronic pain patients. *Pain* 84: 347-352.

Taylor, P.M. (1985) Clinical measurement of pain, distress and discomfort in dogs and cats. In *The Detection and Relief of Pain in Animals – Proceedings of the Animal Welfare Foundation Second Symposium*, ed. Gibson T.E., pp. 76-81. London: BVA Animal Welfare Foundation.

Taylor A.A., Hurnik J.F. and Lehman H. (1995) The application of cost-benefit dominance analysis to the assessment of farm animal quality-of-life. *Social Indicators Research* 35(3): 313-329.

Terwee C.B., Dekker F.W., Wiersinga W.M., Prummel M.F. and Bossuyt P.M.M. (2003) On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. *Quality of Life Research* 12: 349-362.

Theunissen N.C.M., Vogels T.G.C., Koopman G.H.W., Zwinderman K.A.H., Verloove-Vanhorick S.P and Wit J.M. (1998) The proxy problem: child report versus parent report in health-related quality of life research. *Quality of Life Research* 7: 387-397.

Thomas R.J., McEwen J. and Asbury A.J. (1996) The Glasgow Pain Questionnaire – a new generic measure of pain: development and testing. *International Journal of Epidemiology* 25: 1060-1067.

Thurmon, J. C., Tranquilli, W. J. and Benson, G. J. (1996) *Lumb and Jones' Veterinary Anesthesia (3rd Edition)*, pp. 40-41. Baltimore, Md: Williams and Wilkins.

Tugwell P., Wells G., Strand V., Maetzel A., Bombardier C., Crawford B., Dorrier C. and Thompson A. (2000) Clinical improvement as reflected in measures of function and health-related quality of life following treatment with leflunomide compared with methotrexate in patients with rheumatoid arthritis. *Arthritis and Rheumatism* 43(3): 506-514.

Turan A., Karamanlıoğlu B., Memiş D., Usar P, Pamukçu and Türe M. (2004) The analgesic effects of gabapentin after total abdominal hysterectomy. *Anesthesia and Analgesia* 98: 1370-1373.

Turk D.C. (1999) The role of psychological factors in chronic pain. *Acta Anaesthesiologica Scandinavica* 43: 885-888.

Vaillancourt J-P., Martineau G., Morrow M., Marsh W. and Robinson A. (1991) Construction of questionnaires and their use in veterinary medicine. In

Proceedings of the Society of Veterinary Epidemiology and Preventive Medicine, London, pp. 94-106.

Vallerand A.H. (1998) Development and testing of the Inventory of Functional Status-Chronic Pain. *Journal of Pain and Symptom Management* 15(2): 125-133.

Van Dijk M, De Boer JB, Koot HM, Tibboel D., Passchier J and Duivenvoorden H.J (2000). The reliability and validity of the COMFORT scale as a postoperative pain instrument in 0 to 3-year-old infants. *Pain* 84: 367–377.

Varni J.W., Katz E.R., Seid M., Quiggins D.J.L. and Freidman-Bender A. (1998) The Pediatric Cancer Quality of Life Inventory-32 (PCQL-32); I: Reliability and validity. *Cancer* 82(6): 1185-1196.

Varni J.W., Seid M. and Rode C.A. (1999) The PedsQL™: measurement model for the Pediatric Quality of Life Inventory. *Medical Care* 37(2): 126-139.

Varni J.W., Seid M. and Kurtin P.S. (2001) PedsQL™ 4.0: reliability and validity of the Pediatric Quality of Life Inventory™ Version 4.0 generic core scales in healthy and patient populations. *Medical Care* 39(8): 800-812.

Varni J.W., Seid M., Knight T.S., Burwinkle T., Brown J and Szer I.S. (2002a) The PedsQL™ in pediatric oncology. *Arthitis and Rheumatism* 46(3): 714-725.

Varni J.W., Seid M., Knight T.S., Uzark K and Szer I.S. (2002b) The PedsQL™4.0 generic core scales: sensitivity, responsiveness, and impact on clinical decision-making. *Journal of Behavioral Medicine* 25(2): 175-193.

Wade J.B., Price D.D., Hamer R.M., Schwartz S.M. and Hart R.P. (1990) An emotional component analysis of chronic pain. *Pain* 40: 303-310.

Wall P.D. (1976) Editorial. *Pain* 2: 1.

Wall P.D. (1979) On the relation of injury to pain. *Pain* 6: 253-264.

Wall P.D. (1999) Introduction to the fourth edition. In *Textbook of Pain (4th Edition),* eds. Wall P.D. and Melzack R., p.1. Edinburgh: Churchill Livingstone.

Ware J.E. and Sherbourne C.D. (1992) The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Medical Care* 30: 473-483.

Ware J.E. (1995) The status of health assessment 1994. *Annual Review of Public Health* 16: 327-354.

Watson M, Edwards L, Von Essen L, Davidson J., Day R. and Pinkerton R. (1999) Development of the Royal Marsden Paediatric Oncology Quality of Life Questionnaire. *International Journal of Cancer* 83: S65-S70.

Watson P. (2004) Pancreatitis in the dog: dealing with a spectrum of disease. *In Practice* 26(2): 64-77.

Web Centre for Social Research Methods (2004) [online] Available at http://www.socialresearchmethods.net/tutorial/flynn/factor.htm (last accessed 20 February 2005).

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. (1999) The problem of animal subjectivity and its consequences for the scientific measurement of animal suffering. In *Attitudes to Animals*, ed. Dolins F.L., pp. 37-53. Cambridge: Cambridge University Press.

Wemelsfelder, F., Hunter, E.A., Mendl, M.T., Lawrence, A.B. (2000) The spontaneous qualitative assessment of behavioural expressions in pigs: first

explorations of a novel methodology for integrative animal welfare measurement. *Applied Animal Behaviour Science* 67(3): 193-215.

Wemelsfelder F., Hunter T.E., Mendl M.T. and Lawrence A.B. (2001) Assessing the 'whole animal': a free choice profiling approach. *Animal Behaviour* 62: 209-220.

Wemelsfelder F. and Farish M. (2004) Qualitative categories for the interpretation of sheep welfare: a review. *Animal Welfare* 13: 261-268.

Wheeler A.H., Goolkasian P., Baird A.C. and Darden B.V.II (1999) Development of the neck pain and disability scale: item analysis, face and criterion-related validity. *Spine* 24(13): 1290-1294.

WHO (1948) Preamble to the Constitution of the World Health Organization as adopted by the International Health Conference, New York, 19-22 June, 1946; signed on 22 July 1946 by the representatives of 61 States (Official Records of the World Health Organization, no. 2, p. 100) and entered into force on 7 April 1948.

WHOQOL Group. (1995) The World Health Organization quality of life assessment (WHOQOL): position paper from the World Health Organization. *Social Science and Medicine* 41: 1403–1409

Wiebe S., Guyatt G., Weaver B., Matijevic S. and Sidwell C. (2003) Comparative responsiveness of generic and specific quality of life instruments. *Journal of Clinical Epidemiology* 56: 52-60.

Williams A. C.deC., Davies H.T.O. and Chadury Y. (2000) Simple pain rating scales hide complex idiosynchratic meanings. *Pain* 85: 457-463.

Wirnsberger R.M., De Vries J., Jansen T.L.Th.A., Van Heck G.L., Wouters E.F.M. and Drent M. (1999) Impairment of quality of life: rheumatoid arthritis versus sarcoidosis. *The Netherlands Journal of Medicine* 54: 86-95.

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.

Wojciechowska J.I. and Hewson C.J. (2005) Quality-of-life assessment in pet dogs. *Journal of the American Veterinary Medical Association* 226(5): 722-728.

Wolinsky F.D., Wyrwich K.W., Nienaber N.A. and Tierney W.M. (1998) Generic versus disease-specific health status measures. *Evaluation and the Health Professions* 21(2): 216-243.

Woolf C.J. (1994) The dorsal horn: state-dependent sensory processing and the generation of pain. In *Textbook of Pain (3rd Edition),* eds. Wall P.D. and Melzack R., pp. 101-112. Edinburgh: Churchill Livingstone.

Worz R. (2003) Pain in depression – depression in pain. *Pain Clinical Updates (International association for the Study of Pain)* XI(5): 1-4.

Wright J.G. (2003) Interpreting health-related quality of life scores: the simple rule of seven might not be so simple. *Medical Care* 41(5): 597-598.

Wright J.G. and Feinstein A.R. (1992) A comparative contrast of clinimetric and psychometric methods for constructing indexes and rating scales. *Journal of Clinical Epidemiology* 45:1201–1218.

Wright J.G. and Young N.L. (1997) A comparison of different indices of responsiveness. *Jurnal of Clinical Epidemiology* 50(3): 239-246.

Yearley J.H., Hancock D.D. and Mealey K.L. (2004) Survival time, lifespan, and quality of life in dogs with idiopathic Fanconi syndrome. *Journal of the American Veterinary Medical Association* 225(3): 377-383.

Yong H.-H., Bell R., Workman B. and Gibson S.J. (2003) Psychometric properties of the pain attitudes questionnaire (revised) in adult patients with chronic pain. *Pain* 104: 673-681.

Yoxall AT (1978) Pain in small animals – its recognition and control. *Journal of Small Animal Practice* 19: 423-438.

Ziebland S. (1994) Measuring changes in health status. In *Health and Medical Outcomes,* ed. Jenkinson C., pp. 42-53. London: UCL Press.

Zigmond A.S. and Snaith R.P. (1983) The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica* 67: 361-370.

Zimmerman M. (1986) Behavioural investigations of pain in animals. In *Assessing Pain in Farm Animals.* CEC Agriculture Report EUR 9742.

APPENDICES

Please refer to the accompanying CD. Each document on the CD contains all of the appendices for one chapter, with the exception of the content of Appendix 10 (the GUVQuests) which can be found in a separate folder also located on the CD.