BSAVA British Small Animal Veterinary Association

JSAP JOURNAL OF SMALL ANIMAL PRACTICE

PAPER

Development, validation and reliability of a web-based questionnaire to measure health-related quality of life in dogs

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

*School of Veterinary Medicine, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow G12 8QQ

†School of Mathematics and Statistics, College of Science and Engineering, University of Glasgow, Glasgow G12 8QQ

Objectives: To describe the development, preliminary validation and reliability testing of a shortened web-based form of GUVQuest, a structured questionnaire to measure health-related quality of life in dogs.

Methods: The original 109 items were reduced using expert judgement and factor analysis. Validity was established by factor analysis and in a subsequent field trial using a "known groups" approach and classical test theory. Test–retest reliability was assessed using intraclass correlation coefficients.

Results: The instrument comprises 46 items each of which is rated by dog owners using a 7-point Likert scale. Factor analysis revealed a sensible structure containing four health-related quality of life domains (vitality, pain, distress and anxiety) accounting for 64·1% of the variability in the data. The field test involving 125 dogs demonstrated very good discriminative properties and intraclass correlation coefficient values of greater than 0·6.

Clinical Significance: This is the first report of a valid and reliable companion animal health-related quality of life instrument, the contemporary approach to animal welfare measurement, which is presented in a web-based format, with automated production of a health-related quality of life profile. It offers major advantages to dog owners, practitioners and researchers.

Journal of Small Animal Practice (2013) **54**, 227–233
DOI: 10.1111/jsap.12059

Accepted: 25 February 2013; Published online: 5 April 2013

INTRODUCTION

Health-related quality of life (HRQL) is an important patient-reported outcome measure (PROM) in human medicine and structured questionnaires to measure HRQL are developed and tested using well-established psychometric methodology (Streiner & Norman 2008, Abell *et al.* 2009, Brod *et al.* 2009). The gold standard measure of human HRQL is the self report, but where that is impossible (e.g. infants) instruments are completed by an observer who knows the subject well. HRQL instruments can measure the difference between individuals (healthy *versus* sick) or measure change within a patient over time (evaluation of treatment or disease progression). These can be generic or disease-specific, used in a range of contexts, from routine monitoring of wellness to comparing treatment efficacy for particular diseases (Streiner & Norman 1989, Patrick & Erickson 1993, Kane 2006). Such instruments are either profile measures that yield health scores across multiple dimensions or index measures that summarize health status in a single numerical score (Streiner & Norman 1989).

A scientifically sound measurement instrument is valid, reliable and, if designed for evaluative purposes, responsive to clinical change. It should have utility and must be quick and easy to use (Teasdale & Jennett 1974). Validity (criterion, content and construct) provides evidence that the instrument measures what it was designed to measure. Criterion validity is the agreement of a new instrument with some existing "gold standard". Content validity ensures the appropriateness and completeness of the items within the instrument, while construct validity is demonstrated when hypotheses regarding the attribute(s) in question are upheld by use of the instrument. Its reliability reflects the extent to which an instrument can generate the same score when an unchanging subject is measured at two time points, or when two people measure the same subject at one time.

Recently, veterinary scientists have considered the development of instruments to measure HRQL in companion animals (Freeman *et al.* 2005, Yazbek & Fantoni 2005, Brown *et al.* 2007, Budke *et al.* 2008, Hielm-Bjorkman *et al.* 2009, Favrot *et al.* 2010, Niessen *et al.* 2010, Lynch *et al.* 2011, Noli *et al.* 2011, Niessen *et al.* 2012). These generally consist of questions for the companion animal owner, who is well placed to report upon changes in behaviour, attitude and demeanour. The validated paper-based GUVQuest, the first such instrument developed for use in animals, measures chronic pain in dogs through its effect on HRQL (Wiseman-Orr *et al.* 2004). It contains 109 items each of which consists of a descriptor (e.g. active) with a 7-point Likert rating scale, 0 to 6 (with 0 meaning "not at all" and 6 meaning "couldn't be more"). A complex algorithm combines these ratings to create a score profile for the dog, for a number of HRQL domains. GUVQuest has been validated in dogs with painful conditions such as degenerative joint disease (DJD) (Wiseman-Orr 2006), and in non-painful conditions such as lymphoma, where aggressive treatment may impact on a dog's HRQL (unpublished results). In addition, a 60-item version of GUVQuest has shown that HRQL is reduced in obese dogs and that it improves after successful weight loss (German *et al.* 2011). Thus GUVQuest performs as a generic instrument. It is easy to use, but takes 30 minutes to complete, and transformation of responses into the score profile is time-consuming. Accordingly, it was considered important to shorten the instrument and devise an automated data capture and score generation system.

Many instruments developed to measure HRQL in people have been successfully shortened to improve their utility although care must be taken to retain their psychometric properties (Coste *et al.* 1997). Criteria used in selecting items for retention include expert judgement in terms of relevance of items, the use of Factor Analysis (FA) to select items with high loadings, the identification of items that discriminate best, and the use of an item response theory Rasch model or classical test theory (Alcala *et al.* 2004, Osse *et al.* 2007, Las Hayas *et al.* 2010).

The objective of the studies reported here was to produce a shortened generic instrument to measure HRQL in dogs where high utility would facilitate and promote its use by owners and in general practice. This was done by selecting a core of items from GUVQuest, that would demonstrate sensitivity to differences in HRQL of healthy dogs and those suffering diverse disease processes.

MATERIALS AND METHODS

Ethics approval was granted by the University of Glasgow Ethics Committee and written consent was obtained from all clients and participants in the focus group exercise.

Study 1

Statistical focus group exercise

Eight volunteers from the Department of Statistics, University of Glasgow identified items where response distributions differed between disease groups and between well and unwell dogs. One member of the group had a 2nd degree in an unrelated science subject, but seven of the eight participants in the group exercise had more than 4 years statistical experience and a 1st (2) or 2nd (7) degree in statistics. This level of knowledge was considered appropriate for the focus group.

Data were used that had been generated using the 109-item GUVQuest in control dogs belonging to staff at the University of Glasgow Small Animal Hospital (UGSAH) that had undergone a clinical examination to ensure that they were in good health and in clinical cases that were suffering from painful chronic disease or lymphoma. All dogs were allocated a pain score using a 0 to 10 point numerical rating scale (NRS) by the attending clinician. Groups were identified as follows using the hospital diagnosis and the clinician pain score.

- Group 1 – control dogs with a pain score of 0 (n=47);
- Group 2 – dogs with lymphoma with a pain score of 0 (n=41);
- Group 3 – dogs in chronic pain with pain scores of 1 or 2 (n=63);
- Group 4 – dogs in chronic pain with pain scores >6 (n=37).

Using Minitab v.15, the distribution of response scores (0 to 6) for each of the 109 items was presented in standardized panels of four histograms (one for each group) on PowerPoint (Microsoft Office 2003) slides to members of the focus group. Participants were blinded to the nature of the clinical and control groups, but graphs were labelled with the name of the relevant item. Axes were consistent between panels of graphs and across all items.

Each participant worked individually and recorded on a paper proforma whether they considered the shapes of the distributions for each item to be the same or different. A "don't know" response was permitted. Four separate comparisons were made: Group 1 (controls) with Group 2 ("pain-free" lymphoma); Group 1 with Group 3 ("low pain" scores); Group 1 with Group 4 ("high pain" scores) and Group 3 with Group 4.

For each comparison, agreement between seven of eight and eight of eight participants that there was a difference in distributions was taken to mean that there was a difference. Where there was agreement between six of eight and at least one "don't know", these items were further considered by the senior statistician author (EMS) and a decision made as to whether they were showing sufficient difference to be classed as such. Since the focus of this article was discriminative performance of the items, comparisons made were Group 1 (healthy controls) with Group 2 (unwell, but no pain) and Group 1 with Group 3 (unwell with low pain), on the basis that detecting a difference between these groups would require more sensitivity than differentiating between healthy controls and Group 4 (unwell with high pain). Comparison of Group 3 with Group 4 was carried out; its further analysis will contribute to the future development of an evaluative rather than a discriminative instrument.

Factor Analysis (FA)

Items identified in the group exercise as those that distinguished between Group 1 and Groups 2 and 3 were extracted from the original 109 items and FA was carried out (Minitab v.15). A principal components method of FA with a varimax rotation was performed. Input variables were all item ratings. Loadings were sorted, and items with loadings of less than 0·3 were excluded. Guided by a scree test and the Kaiser criterion the interpretability of a range of factor models was examined. Factors were interpreted based on how those items loading onto a particular factor were related (and unrelated to items not loading onto that factor). A factor model was sought that accounted for an acceptable amount of the variability in the data, was readily interpretable, and did not include any factors containing only one or two items.

Development of a web-based instrument

Software developers Kyria Ltd. (http://kyria.co.uk/) were commissioned to develop a web-based instrument using the reduced number of items and their associated algorithm. A detailed brief was provided which, along with staged development and pretesting, was designed to ensure that utility and functionality would be optimized.

Study 2

Field test

Over a 15-month period, owners of unwell dogs attending GUSAH or selected VetsNow emergency clinics, and owners of dogs recruited from staff at GUSAH, deemed to be free of disease by author JR, completed at least one online questionnaire. No restrictions, apart from acute trauma in the case of dogs admitted to the emergency clinics, were imposed on the inclusion of unwell dogs. Owners of unwell dogs were recruited from the daily case load by a senior nurse in GUSAH or equivalent in VetsNow clinics as and when it was logistically possible. Accordingly, the sampling was best described as cluster.

Scores generation and data analysis

A score for each factor identified in FA (domain of HRQL) was obtained by calculating the mean rating for all items loading onto that factor. Scores for all domains provide a score profile for each dog. This profile of scores in four domains of HRQL (see results below) for each dog was generated automatically by the system software and stored in an Excel (Microsoft Office 2003) spreadsheet. Descriptive statistics were used to identify differences between the well and unwell groups, and formal statistical analysis using nonparametric Mann-Whitney tests due to the non-normality of the data followed. Test–retest reliability was assessed using the intraclass correlation coefficient (ICC). A one-way random model was assumed where the subjects are assumed random (Shrout & Fleiss 1979).

RESULTS

Study 1

Identification of differences

Figure 1 illustrates examples of complete agreement in the focus group that there was a difference between control and low pain groups for the item athletic (Fig 1a) and complete agreement that there was no difference between control and low pain groups for the item alert (Fig 1b).

Following comparison of Group 1 (control) with Group 2 (pain-free lymphoma) eight of eight and seven of the eight experts agreed that 8 items and 23 items, respectively, distinguished between the groups. Similarly, following comparison of Group 1 with Group 3 (low pain), eight of eight and seven of the eight experts agreed that 8 items and 12 items, respectively, distinguished between the groups. On that basis, 31 items from the comparison between Groups 1 and 2 and 20 items from the comparison between Groups 1 and 3 were selected for inclusion in the generic questionnaire.

When Group 1 was compared with Group 2, seven items were considered borderline on the basis that six of the eight experts considered that there was a difference between the groups, one of eight considered them to be the same and one of eight recorded a

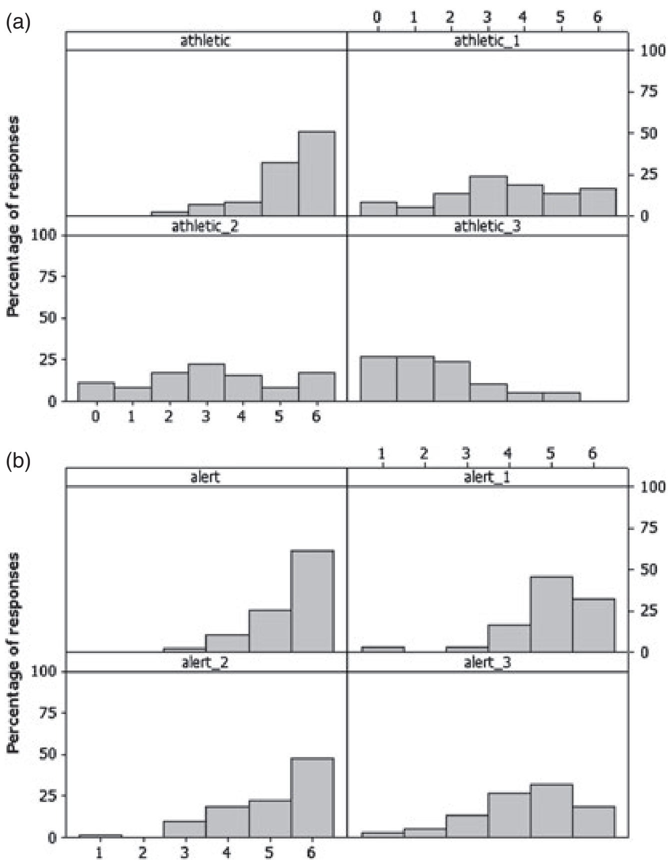

FIG 1. Panel of histograms of owner responses for (a) the item athletic and (b) the item alert for control (top left), lymphoma (top right), high pain (bottom right) and low pain (bottom left) dogs, as presented to the focus group (eight individuals) in PowerPoint format

"don't know" response. Of these seven items, two achieved a score of seven of eight or eight of eight in the control *versus* low pain group and so were selected on that basis. The remaining five were reassessed by EMS and one was considered not to distinguish between the groups and was rejected. The other four items were classified as showing a difference. Similarly, using the same criteria, when Groups 1 and 3 were compared, 14 items were borderline. Of these 14, 7 had a score of seven of eight or eight of eight in the control *versus* lymphoma group and so were selected on that basis. The remaining seven items were reassessed as before. Following reassessment, all seven were classified as showing a difference and were included in the generic questionnaire (Fig 2a, b). Thirteen items were common to both group comparisons. Table 1 shows examples of items selected as a result of comparison of Group 1 with Groups 2 and 3. When Group 3 was compared with Group 4, 25 items showed a difference between the groups. Of these, 14 showed evaluative but not discriminative potential, while the other 11 showed both.

Factor Analysis (FA)

FA was performed on responses for the 49 items, obtained from 215 questionnaires, each one completed at the time of first presentation by owners of dogs referred to GUSAH with a diagnosis of lymphoma (n=49) or chronic painful disease (n=166), of which 107 were suffering from DJD. A scree plot and the Kaiser criterion suggested that a model containing approximately four factors was most appropriate. Examination of the items loading onto each factor and the consequent interpretability was considered for a range of models, taking into consideration the amount of variability in data that was accounted for by each of the factor models. A 4-factor model, accounting for 60% of the variance and consisting of factors that could be interpreted as "vitality", "pain", "distress" and "anxiety" was most suitable. Items "greedy" and "stretching" did not load onto any of the four factors and had low communalities, so were removed. FA of the remaining 47 items produced a 4-factor model that accounted for 62% of the variance, but "stoical" no longer loaded on to any factor, had low communality and was removed. Final FA of the remaining 46 items resulted in a 4-factor model that accounted for 64·1% of the variance, with items loading onto multiple factors as follows: vitality (34), pain (17), distress (12) and anxiety (13).

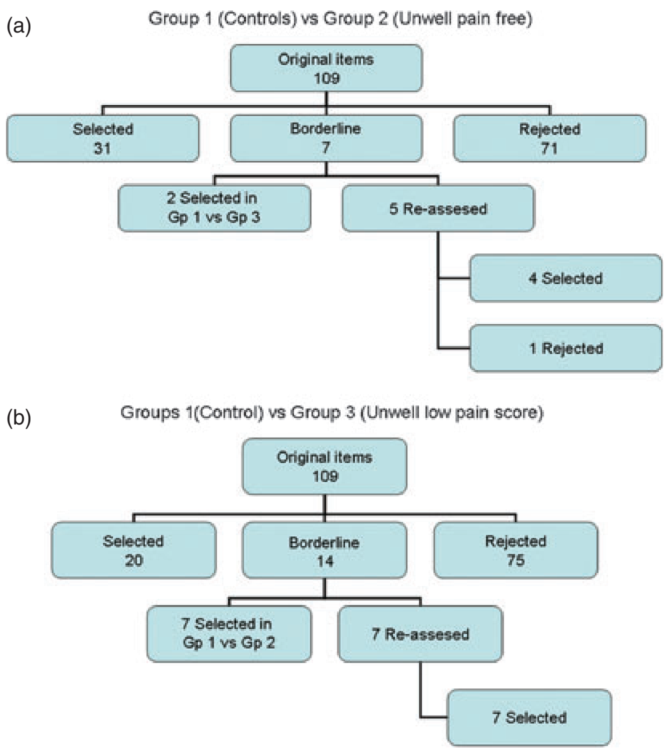

FIG 2. Flow diagrams showing the process of item selection when (a) Group 1 (pain-free controls) was compared with Group 2 (unwell pain-free) and (b) when Group 1 was compared with Group 3 (unwell with low pain scores)

Table 1 Examples of items showing differences between Groups 1 and 2 and Groups 1 and 3

Items showing differences between controls (G1) vs lymphoma (G2)	Items showing differences between controls (G1) vs low pain (G3)
Active	
Alert	
Apathetic	
Athletic	Athletic
Bright	
Calm	
Confused	
	Complaining
	Confident
	Consistent
Depressed	Depressed
	Energetic
Fit	Fit
Funloving	
Greedy	
Groaning	Groaning
Happy	Happy
	Limping
	Pained
	Relaxed

Development of a web-based instrument

A 46-item web-based instrument was pretested which led to minor revisions to the design and content. Generally, owners completed the questionnaire in 5 minutes or less. An automated system for emailing a username and password to new users and for sending automatic reminders to complete the web-based instrument, was integrated and tested to ensure its utility.

Study 2

Thirty-five healthy control dogs comprising 15 females and 20 males with a mean age of 5·0 ±3·0 years (range 1 to 11) and 90 unwell dogs of which there were 48 females and 42 males with a mean age of 6·7 ±4·0 years (range 1 to 18), representing a comprehensive range of breeds were recruited. Unwell dogs were presented to a range of disciplines (Table 2). Differences between well and unwell dogs existed for all four factors, and there was greater variability in the unwell group compared with the well group (Fig 3). The instrument demonstrated very good discriminative properties (Table 3).

Sixteen owners of control dogs completed two questionnaires, 2 weeks apart. The differences in scores between questionnaires

Table 2. Initial presenting disciplines for unwell dogs

Discipline	Number of cases
Radiotherapy	1
Cardiology	2
Pain clinic	4
Physiotherapy	7
Neurology	9
Soft tissue surgery	9
Ophthalmology	10
Oncology	11
Orthopaedics	14
Internal medicine	23

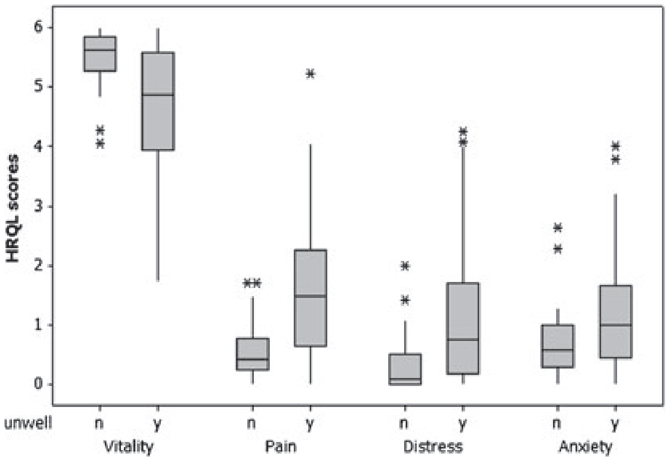

FIG 3. Plots of scores for four domains of HRQL (vitality, pain, distress and anxiety) generated by the owners of 35 healthy control dogs (N) and 90 unwell dogs (Y) using the 46 item web-based HRQL instrument

Table 3. Mann-Whitney test results comparing the scores of healthy and unwell dogs for each of the four domains (vitality, pain, distress and anxiety)

	Difference in median (well–unwell)	P-value	95% confidence interval
Vitality	0·65	<0·001	(0·32, 1·06)
Pain	–0·94	<0·001	(–1·35, –0·53)
Distress	–0·5	<0·001	(–0·83, –0·25)
Anxiety	–0·35	<0·006	(–0·64, –0·07)

1 and 2 are depicted in Fig 4. The ICC (95% confidence intervals) for the three domains excluding distress (where the majority of values were 0) were vitality 0·867 (0·67 to 0·95), pain 0·655 (0·25 to 0·87) and anxiety 0·608 (0·19 to 0·84).

DISCUSSION

While others have developed single index instruments to assess orthopaedic pain (Brown *et al.* 2007, Hielm-Bjorkman *et al.* 2009), GUVQuest was the first psychometric instrument to measure chronic pain in dogs by assessing its impact on quality of life. It generates a profile of scores and while a single measure of overall health makes it easy to ascertain whether the net effect of an intervention is positive or negative (Brazier *et al.* 2007), it may be less sensitive to differences between groups and to changes in health over time (Streiner & Norman 1989). This was borne out by Mutebi *et al.* (2011) who demonstrated that moving from the Medical Outcomes Study 36-Item Short Form Health Survey (SF-36) profile measure to the SF-6D, a generic single index measure derived from the SF-36, resulted in decreased discriminative and evaluative properties. Accordingly, it was considered important that a shortened version of the GUVQuest should retain the capability to generate a profile of scores.

For an instrument to be adopted it must have utility (Teasdale & Jennett 1974). The original GUVQuest took 30 minutes to complete and its complex scoring algorithm required a time-consuming calculation. The shortened web-based instrument, a structured questionnaire consisting of 46 items which generally takes 5 minutes or less to complete and with automatic and instantaneous generation of scores, demonstrates greatly improved utility.

The shortening of GUVQuest, used expert judgement and factor loadings, both widely used approaches (Alcala *et al.* 2004, Osse *et al.* 2007, Las Hayas *et al.* 2010). The latter shortened the Health-Related Quality of Life for eating disorders questionnaire (HeRQoLEDv2) using Rasch analysis and concluded that the short form maintained good psychometric properties. Ware

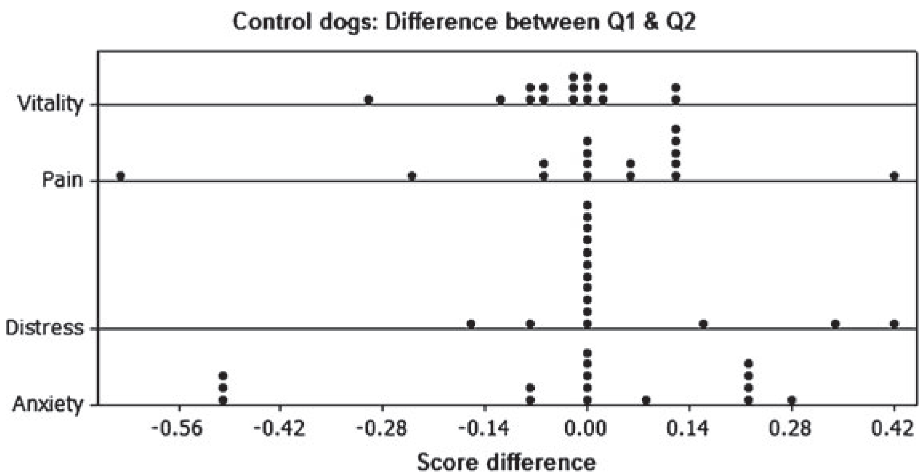

FIG 4. Differences in scores for four domains of HRQL (vitality, pain, distress and anxiety) generated by 16 owners of healthy dogs 2 weeks apart. Q1, Questionnaire 1; Q2, Questionnaire 2

et al. (1996) used regression methods to select 12 of the 36 items in the SF-36 to produce the 12-Item Short Form Health Survey (SF-12). Completion time was cut by one third and while some information was lost, several studies have indicated very good correlation and agreement (Failde *et al*. 2009, 2010).

The most important property of an instrument is its validity. In the case of dog HRQL measurement, no "gold standard" exists so criterion validity could not be evaluated. The content validity of the GUVQuest was established during the process of its development (Wiseman-Orr *et al*. 2004) and the robust process of item reduction contributes to the content validity of the short form. Construct validity can be assessed in a variety of ways. Factorial validity requires the statistical analysis of correlations between responses to the items of an instrument to determine if an underlying factor structure fits the construct upon which the instrument was developed (Johnston 1998). A good factor model is one in which the derived factors are interpretable (Johnston 1998) as was the case in our study. Although a larger number of factors will account for more of the variance, a smaller number of factors is more manageable. FA revealed a sensible 4-factor model that accounted for 64% of the variability in the data. By comparison a 4-factor quality of life questionnaire regarding infants (Manificat *et al*. 1999) accounted for 45% of the variance, and an 11-factor questionnaire designed to measure the behaviour and temperament of pet dogs (Hsu & Serpell 2003) accounted for 57% of the variance. Accordingly, the percentage of variance accounted for by the 4-factor model obtained from our study is within the range of that reported for proxy human HRQL questionnaire instruments and owner completed questionnaires for dogs. In addition, the field study upheld the hypothesis that the instrument would discriminate between the HRQL of healthy dogs compared with those that were unwell. An extension of this work using larger group sizes will be reported in a subsequent publication.

The ICC values for the domains were greater than 0·6, indicating that test–retest reliability conducted with a 2-week interval for the web instrument was good (Rosner 2005). It was assumed that the health status of control dogs would not change over the 2-week period between the completion of questionnaires, and respondents would not remember their previous responses.

The importance of electronic instruments for human health measurement is widely recognized (WHO 2006). Comparison of web-based and paper-based pain assessment instruments for people has shown similar results, and a respondent preference for the web-based version (Cook *et al*. 2004). Furthermore, a multi-centre study of an electronic data capture system for people with rheumatoid arthritis indicated that access to health status information promotes patient satisfaction and patient–clinician interactions (Huffstutter *et al*. 2007). This may be highly relevant to clinicians seeking to promote owner compliance in canine chronic disease management programmes. Consequently, it was considered important to make the shortened instrument available online to improve its utility and obtain such benefits. A web-based instrument allows the respondent to complete the questionnaire in the dog's normal environment, and it facilitates automatic data capture, computation of scores and simultaneous reporting of results. It may also facilitate better protocol compliance (Palmblad & Tiplady 2004).

This novel web-based instrument described here could be used routinely in general practice to facilitate the management of chronic disease and owner engagement (Yeates & Main 2009) and in clinical trials and research it is likely to offer significant advantages over current methods.

Acknowledgements

The authors wish to thank all the dog owners who willingly participated in our studies as well as the statistical volunteers, nursing staff and clinicians at the University of Glasgow Small Animal Hospital and VetsNow emergency clinics.

Conflict of interest

Authors Reid and Wiseman-Orr are Directors in NewMetrica Ltd., a company specializing in the development of instruments to measure pain and quality of life in non-verbal populations.

References

Abell, N., Springer, D.W. & Kamata, A. (2009) Developing and Validating Rapid Assessment Instruments. New York, NY, USA: Oxford University Press

Alcala M. J., Casellas F., Fontanet G., *et al*. (2004). Shortened questionnaire on quality of life for inflammatory bowel disease. *Inflammatory Bowel Disease* **10**, 383-391

Brazier, J. E., Ratcliffe, J., Tsuchiya, A., *et al*. (2007). Measuring and Valueing Health for Economic Evaluation. Oxford University Press, Oxford, UK

Brod, M., Tesler, L. E. & Christensen, T. L. (2009) Qualitative research and content validity: developing best practices based on science and experience. *Quality of Life Research* **18**, 1263-1278

Brown, D. C., Boston, R. C., Coyne, J. C., *et al*. (2007) Development and psychometric testing of an instrument designed to measure chronic pain in dogs with osteoarthritis. *American Journal of Veterinary Research* **68**, 631-637

Budke, C. M., Levine, J. M., Kerwin, S. C., *et al*. (2008) Evaluation of a questionnaire for obtaining owner-perceived, weighted quality-of-life assessments for dogs with spinal cord injuries. *Journal of the American Veterinary Medical Association* **233**, 925-930

Cook, A. J., Roberts, D. A., Henderson, M. D., *et al*. (2004) Electronic pain questionnaires: a randomised crossover comparison with paper questionnaires for chronic pain assessment. *Pain* **110**, 310-317

Coste, J., Guillemin, F., Pouchot, J., *et al*. (1997) Methodological approaches to shortening composite measurement scales. *Journal of Clinical Epidemiology* **50**, 247-252

Failde, I., Medina, P., Ramirez, C.(2009) Assessing health related quality of life among coronary patients : SF - 36 vs SF - 12. *Public Health* **123**, 615-617

Failde, I., Medina, P., Ramirez, C., *et al*. (2010) Construct and criterion validity of the SF - 12 health questionnaire in patients with acute myocardial infarction and unstable angina. *Journal of Evaluation in Clinical Practice* **3**, 569-573

Favrot, C., Linek, M., Mueller, R., *et al*. (2010) Development of a questionnaire to assess the impact of atopic dermatitis on health-related quality of life of affected dogs and their owners. *Veterinary Dermatology* **21**, 63-69

Freeman, L. M., Rush, J. E., Farabaugh, A. E., *et al*. (2005) Development and evaluation of a questionnaire for assessing health-related quality of life in dogs with cardiac disease. *Journal of American Veterinary Medical Association* **226**, 1864-1868

German, A. J., Holden, S. L., Wiseman-Orr, M. L., *et al*. (2012) Quality of life is reduced in obese dogs but improves after successful weight loss. *Veterinary Journal* **192**, 428-434

Hielm-Bjorkman, A. K., Rita, H. & Tulamo, R.M. (2009) Psychometric testing of the Helsinki chronic pain index by completion of a questionnaire in Finnish by owners of dogs with chronic signs of pain caused by osteoarthritis. *American Journal of Veterinary Research* **70**, 727-734

Hsu, Y. & Serpell, J. A. (2003) Development and validation of a questionnaire for measuring behavior and temperament traits in pet dogs.*Journal of American Veterinary Medical Association* **223**, 1293-1300

Huffstutter, J., Craig, W. D., Schimizzi, G., *et al*. (2007) A multicenter, randomized, open study to evaluate the impact of an electronic data capture system on the care of patients with rheumatoid arthritis. *Current Medical Research and Opinion* **23**, 1967-1979

Hurst, N. P., Ruta, D. A. & Kind, P. (1998). Comparison of the MOS Short Form-12 (SF12) health status questionnaire with the SF36 in patients with rheumatoid arthritis. *British Journal of Rheumatology* **37**, 862-869

Johnston, C. C. (1998) Psychometric issues in the measurement of pain. In Measurement of Pain in Infants and Children. Eds G. A. Finley, P. J. McGrath. *Progress in Pain Research and Management*. **10**, IASP Press, Seattle, USA. 5-20

Kane, L. R. (2006). Understanding Health Care Outcomes Research. 2nd edn. Jones and Bartlet, Sudbury, Massachusetts, MA, USA

Las Hayas, C., Quintana, J. M., Padierna, J. A., *et al*. (2010) Use of rasch methodology to develop a short version of the health related quality of life for eating disorders questionnaire: a prospective study. *Health and Quality of Life outcomes* **8**, 29

Lynch, S., Savary-Bataille, K., Leeuw, B., *et al*. (2011) Development of a questionnaire assessing health-related quality-of-life in dogs and cats with cancer. *Veterinary and Comparative Oncology* **9**, 172-182

Manificat, S., Dazord, A., Langue, J., *et al*. (1999) A new instrument to evaluate infant quality of life. *MAPI Research Institute Quality of Life Newsletter* **23**, 7-8

Mutebi, A., Brazier, J. E. & Walters, S. J. (2011) A comparison of the discriminative and evaluative properties of the SF-36 and the SF-6D index. *Quality of Life Research* **20**, 1477-1486

Niessen, S. J, Powney, S, Guitian, J, *et al*. (2010) Evaluation of a quality of life tool for cats with diabetes Mellitus. *Journal of Veterinary Internal Medicine* **24**, 1098-1105

Niessen, S. J., Powney, S., Guitian, J., *et al*. (2012) Evaluation of a quality of life tool for dogs with diabetes Mellitus. *Journal of Veterinary Internal Medicine* **26**, 953-961

Noli, C., Minafo, G., Galzerano, M. (2011) Quality of life of dogs with skin diseases and their owners. Part 1: development and validation of a questionnaire. *Veterinary Dermatology* **22**, 335-343

Osse, B. H. P., Vernooij-Dassen, M., Schade, E., *et al*. (2007) A practical instrument to explore patients' needs in palliative care: the problems and needs in palliative care questionnaire- short version. *Palliative Medicine* **21**, 391-399

Palmblad, M. & Tiplady, B. (2004) Electronic diaries and questionnaires: designing user interfaces that are easy for all patients to use. *Quality of Life Research* **13**, 1199-1207

Patrick, D. L. & Erickson, P. (1993). Health Status and Health Policy: Quality of Life in Health Care Evaluation and Resource Allocation. Oxford University Press, New York, NY, USA

Rosner, B. (2005) Fundamentals of Biostatistics. Duxbury Press, Belmont, CA, USA

Shrout, P. E. & Fleiss, J. L. (1979) Intraclass correlation: uses in assessing rater reliability. *Psychological Bulletin* **86**, 420-428

Streiner, D. L. & Norman, G. R. (2008) Health Measurement Scales: A Practical Guide to Their Development and Use. 4th edn. Oxford University Press, Oxford, UK

Teasdale, G. & Jennett, B.(1974) Assessment of coma and impaired consciousness. A practical scale. *Lancet* **2**, 81-84

Ware, J. Jr, Kosinski, M. & Keller, S. D. (1996) A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. *Medical Care* **34**, 220-233

Wiseman-Orr, M. L., Nolan, A. M., Reid, J., *et al*. (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**, 1077-1084

Wiseman-Orr, M. L., Scott, E. M., Reid, J., *et al*. (2006) Validation of a structured questionnaire as an instrument to measure chronic pain in dogs on the basis of effects on health-related quality of life. *American Journal of Veterinary Research* **67**, 1826-1836

World Health Organisation. (2006) eHealth Tools & Services: needs of the member states. Report of the WHO global observatory for ehealth

Yazbek, K. V. & Fantoni, D. T. (2005) Validity of a health-related quality-of-life scale for dogs with signs of pain secondary to cancer. *Journal of American Veterinary Medical Association* **226**, 1354-1358

Yeates, J. & Main, D, (2009) Assessment of companion animal quality of life in veterinary practice and research. *Journal of Small Animal Practice* **50**, 274-281