Abstract

Objective: Development, initial validation and reliability testing of a shortened version of a web-based questionnaire instrument to measure generic health-related quality of life (HRQL) in companion dogs, to facilitate smartphone as well as online delivery.

Methods: The original 46 items were reduced using expert judgement and factor analysis (FA). Items were removed on the basis of item loadings and communalities on factors identified through FA of responses from owners of healthy and unwell dogs, intra-factor item correlations, readability of items in the UK, USA and Australia and ability of individual items to discriminate between healthy and unwell dogs. Some evidence for validity was established by FA and in a subsequent field trial using a “known groups” approach. Test–retest reliability was assessed using intraclass correlation coefficients (ICC).

Results: The instrument comprises 22 items, each of which is rated by dog owners using a 7 point Likert scale. Factor analysis revealed a structure with four HRQL domains (Energetic/Enthusiastic, Happy/Content, Active/Comfortable, and Calm/Relaxed) accounting for 72% of the variability in the data compared with 64% for the original instrument. The field test involving 153 healthy and unwell dogs demonstrated very good discriminative properties (15% misclassification) and ICC values of greater than 0—7.

Conclusions and Clinical Relevance: The 22 item shortened form possesses improved measurement properties compared with the original instrument and can be accessed via a mobile phone app. This is likely to increase the acceptability to dog owners of the use of the instrument for assessment of HRQL in veterinary practice as a routine wellness measure in healthcare packages and as a therapeutic monitoring tool.

Keywords : Canine, Quality of Life, measurement, smartphone app

Introduction

The measurement of health-related quality of life (HRQL) plays an increasingly important part in human medicine to detect disease (discriminative purposes) and to measure change in health status over time (evaluative purposes) (Fayers & Machin, 2013). Structured questionnaire instruments to measure HRQL of people are developed and tested using well established psychometric methodology (Streiner & Norman, 2008; Abell et al., 2009; Brod et al.2009). These instruments are designed for self-report by the subject, but where self-report is not possible (e.g. infants and the cognitively impaired) they are completed by an observer who knows the subject well. Instruments can be disease specific, focusing on a particular condition, or they can be generic, designed to be used in a variety of circumstances. Instruments to measure HRQL in companion animals generally consist of questions for the pet owner, who is well placed to report upon the often subtle changes in behaviour, attitude and demeanour that occur with chronic disease. The majority of HRQL instruments that have been developed for companion animals are disease specific [Freeman et al., 2005 (cardiac disease); Yazbek & Fantoni, 2005 (cancer); Budke et al., 2008 (spinal cord injuries); Favrot et al., 2010 (atopic dermatitis); Lynch et al., 2011 (cancer); Noli et al., 2011 (skin disease); Niessen et al. 2012 (diabetes mellitus. However, this paper describes the shortening of a 46 item generic instrument (VetMetrica, www.newmetrica.com) which measures the impact of chronic pain and non-painful physical chronic diseases on the dog’s quality of life (xxxx). Disease-specific instruments may be more responsive to clinical change, but generic instruments can be valuable indicators of a range of impacts associated with disease and its treatment, and may be the only option when a patient is suffering from more than one condition, as is often the case in older companion animals. A substantial use for the instrument exists within the veterinary practice, including raising the profile of preventative veterinary medicine within a health and wellness model, where regular use of the instrument between routine visits for vaccination enhances communication with clients and establishes stronger bonds with them as partners in their animals' healthcare. Furthermore it improves disease detection, including chronic disease which is often unrecognised and unreported, The validity of the instrument is currently being tested for the purposes of measuring clinical change in response to treatment for a variety of chronic conditions and initial results are encouraging (Yam et al., 2016). Users report that being able confidently to demonstrate a deteriorating QOL to owners would help to facilitate end-of-life decision-making for individual dogs. Other than the subject of this paper, only 2 generic HRQL instruments to measure HRQL in dogs have been published (Wojciechowska et al., 2005; Lavan, 2013), but one was shown not to distinguish healthy from sick dogs (Wojciechowska et al., 2005) and the use of the other (Lavan, 2013) is restricted to use in healthy dogs.

Instrument development is an iterative process, in which instruments are refined and re-tested with new populations in new contexts, and instruments developed to measure HRQL in people have been shortened to improve their usefulness, for example shortening of the generic SF36 item HRQL instrument for people to the SF12 item version (Cheak-Zamora, 2009). Guidelines for shortening existing composite scales such as those designed to measure HRQL are scarce. In 1997, Coste et al. reported that the process of scale shortening lacked rigorous methodology and this was confirmed by Stanton et al (2002). Criteria used previously to select items for retention include expert judgement (Osse et al., 2007), the identification of items that discriminate best (Reid et al., 2013) and the use of statistical techniques such as Factor Analysis (FA) (Las Hayas et al. 2010; Reid et al., 2013). Factor Analysis is a statistical technique that analyses the relationships between variables, in this case questionnaire item responses, and clusters them into a small number of homogenous groups which can then be used for analysis. Groupings of variables revealed by such analysis, which in respect of healthcare are also related on clinical or other grounds, are termed factors and the association between a variable and factor is expressed as a factor loading of the variable (values between 0 and 1), where the higher the loading the closer the association. The communality of a variable is the portion of the variance of that variable that is accounted for by the common factors (DeVellis, 2011). Although FA is capable of providing any number of factor models for a given data set, there are established methods which can be used to identify how many factors could sensibly be extracted, including the scree test and the Kaiser criterion (Coste et al., 2005). However, it is up to the instrument developer to decide upon the most satisfactory factor model, the number of factors it contains, and name these according to the interpretation of their associated items. Importantly, a good factor model is one in which the derived factors are readily interpretable and which accounts for a reasonable amount of the variance in the data set from which it was created (StatSoft Inc., 2003). With current software programs, it is possible to rapidly perform FA with various values for the number of factors to be extracted and select the model that is most sensible on clinical or other grounds (StatSoft. Inc., 2003) The most common type of factor analysis is Exploratory Factor Analysis (EFA) where the factor loadings of each item are used to determine the factor structure of a data set collected for the purpose of instrument development. Confirmatory Factor Analysis (CFA) determines whether analysis of a new data set performs in the same way with items loading, as predicted, on the expected number of factors, thus testing the validity of the factor solutions obtained from the EFA (Floyd & Widaman, 1995).

In addition to its use in instrument development and shortening, FA is one of the most commonly used procedures in the validation of psychological measures (Nunnally & Bernstein, 1994; Floyd & Widaman, 1995). 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 and is established during its construction (Fayers & Machin, 2013). A number of approaches exist to examine the construct validity of a new questionnaire, and these include factorial validity and known-groups validity. Factorial validity is demonstrated if, after FA, an interpretable factor structure fits the construct that the instrument was designed to measure (Johnston, 1998), a construct being something that is not directly observable or measurable, like happiness for example. In the context of this paper the construct being measured was HRQL which is the subjective evaluation by an individual of its circumstances that include an altered health state and related interventions (Wiseman Orr et al., 2006). In the known groups approach to determine the construct validity of an instrument, predictions are made about how scores obtained with the instrument will differ between groups and these predictions are then tested. For example, an instrument should be able to distinguish correctly between groups that would be expected to have quite different scores, such as healthy and unwell animals (Reid et al., 2013). In addition to testing an instrument's validity, evidence should also be sought for its reliability which is necessary (although not sufficient) for validity. A reliable instrument is one that will produce the same score when an unchanging subject is measured at two time points by the same observer (repeatability/intra-rater reliability), or when two people measure the same subject at one time (reproducibility/inter-rater reliability) (Streiner & Norman, 2008).

In addition to validity and reliability, to be useful in a clinical setting an instrument must also have utility - it must be acceptably quick, easy to understand and simple to use (Teasdale & Jennett, 1974). In terms of speed and ease of use, electronic technologies have offered much promise for health assessment including providing an acceptable, and in many cases a preferable, alternative to paper, regardless of user age and previous experience of computers (Greenwood et al., 2006). Access to such assessment instruments may be via the web, and preferably in a form that is compatible with mobile platforms. The use of smartphones has revolutionised the communication landscape, providing real-time, on-demand communication and more flexibility compared with other mobile technologies (Boulos et al., 2014). In human healthcare, mobile health (m-health) applications are increasing, with many clinicians and allied health workers already adopting smartphones successfully in a diverse range of practices (Boulos et al, 2014). However, these apps must be carefully designed to retain the utility of the instrument they deliver. Krebs et al (2015) surveyed US health app users and ‘too much time to enter data’ with consequent loss of interest was reported by 44.5% respondents as the reason why they had discontinued use of the app. There appears to be potential animal health and welfare benefits to developing an app which would provide robust measurement of canine HRQL as part of a veterinary care package. An instrument for providing canine HRQL measurement has been developed and evaluated (xxxx xxxx xxxx xxxx) however from a practical perspective, the resulting 46 items were considered too many for presentation via an app. This paper reports the process undertaken to shorten the 46 item instrument and to investigate the measurement properties of the short form to facilitate development of an app designed for owner completion as part of a care package provided by veterinary surgeons.

Materials & Methods

Original instrument

A 46 item long form web-based questionnaire instrument to measure canine HRQL had been developed from an original, novel, paper-based canine HRQL instrument, and both the original and web-based version had previously been validated (xxxx; xxxx). This generic instrument consisted of 46 questions for the dog owner, each of which comprised a descriptor (eg ‘active’) with a 7-point Likert rating scale, 0 – 6 (with 0 meaning ‘not at all’ and 6 meaning ‘couldn’t be more’. During development of the 46 item instrument owner responses to the items were used to generate an HRQL profile comprising scores in 4 QOL domains – named by the instrument developers as vitality, pain, distress and anxiety in accordance with the items loading onto them. The 46 item instrument was shown to have high utility, was easy to use, taking around 5 minutes to complete online, and with automatic and instantaneous transformation of responses into the scores profile.

Confirmatory factor analysis

Previously, as part of a field test to determine the known groups validity and reliability of the 46 item questionnaire instrument, owners of unwell dogs attending xxxxx and selected Vets Now clinics and owners of dogs recruited from clinical, non clinical, nursing and administrative staff members at xxxx deemed to be generally healthy by author xx on the basis of history and lack of clinical signs completed at least 1 online assessment using the 46 item instrument between January 2011 and April 2012 (xxxx). The only inclusion criterion for unwell dogs was that the dog was suffering from a non-acute condition that was expected to affect its QOL. Owners of unwell dogs were recruited from the daily case load by a senior nurse in xxxxx as and when it was logistically possible with no attempt made to control selection bias. Accordingly the sampling was best described as cluster. Ethics approval was granted by the xxxxxxx and written consent was obtained from all owners.

Factor analysis was carried out (Minitab v.16) on the first questionnaire completed by each owner. 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 <0.3 were excluded (Floyd & Widaman, 1995). Guided by a scree test and the Kaiser criterion, the interpretability of a range of factor models was examined. 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 1 or 2 items (Norman & Streiner, 1994).

Reduction of items

Using the results of the CFA, any item with a loading onto any factor <0.5 was removed (Shevlin & Miles, 1998). Thereafter any item with a communality of <0.5 was removed (Velicer et al., 1982; MacCallum et al., 1999). The remaining items were considered for removal on the basis of their correlation with other domain items, readability and ability to discriminate between healthy and unwell dogs. A Pearson Coefficient was calculated for each item and those with a correlation of ≥0.80 or ≤0.20 with other item(s) in the same domain were considered for removal on the basis that they were too similar to others in the domain and therefore extraneous (≥ 0.80) or not related to the underlying construct of the domain (≤ 0.20) (Boyle, 1991; Coste et al., 1995). To ensure that the instrument could be used in non-UK English-speaking countries, in which some words might have slightly different meanings and common uses than in the UK, the suitability of the items was tested by means of two small surveys, one in USA (n = 9) and one in Australia (n = 15). These asked adult respondents to identify items that they considered not to relate to dogs: respondents were dog owners identified and contacted via email by authors xxxx or by veterinary surgeons abroad. In addition, in the absence of established readability metrics for individual words (rather than continuous prose), a number of novel approaches were used to test that items would be readily understood by most adults in the UK. For example, items were reviewed by a class of 9-year old schoolchildren and by a group of adult literacy tutors, and their inclusion in two dictionaries for children aged 9-12 years was checked. All decisions to remove items were made primarily on the basis of the surveys in USA and in Australia, but the UK studies showed that many of those items would also cause difficulty to some UK readers, and showed that none of the remaining items would cause such difficulty. To identify items that could discriminate well between unwell dogs and healthy dogs, histograms of item responses for each item for healthy and for unwell dogs, each plotted on a single graph, were constructed, and those considered by the authors not to discriminate well, on the grounds that the two histograms looked very similar, were removed. Initial screening was carried out by 2 authors (xx and xx) and where there was disagreement regarding discrimination a final judgement was made by the third author, a statistical expert.

Factor analysis of the items retained for the shortened instrument

Using the same data set as was used for the CFA and selection of items, those items retained for the shortened instrument were extracted and subjected to FA as for the CFA, with the exception that items with loadings of <0.5 were excluded. Two, 3 and 4 factor models were explored to determine the optimum factor structure for the shortened instrument and an algorithm, based on the item–factor associations of the selected factor model, was derived in order to generate a domain score for each of the resultant factors/domains.

Field test of the shortened instrument

A different group of owners of healthy dogs, and owners of unwell dogs recruited at xxxxxxxxxx and according to the clinical judgement of 3 vets in general practice and a pharmaceutical veterinary advisor completed at least 1 online assessment using the shortened instrument between February and April 2014. No attempt was made to control selection bias, but each owner was confirmed as the primary carer of the dog. Using the first assessment completed for each dog, descriptive statistics were used to identify differences between the healthy and unwell groups, and this was followed by formal statistical analysis using non-parametric Mann-Whitney tests due to the non-normality of the data. Linear discriminant analysis was carried out to determine the accuracy of the instrument in differentiating healthy from unwell dogs. The same owner of a number of healthy dogs completed 2 assessments, 2 weeks apart, and 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

Confirmatory factor analysis

Factor analysis was carried out on responses from owners of 88 unwell dogs and 34 healthy dogs (Table 1) who participated in the field test to determine the validity of the original 46 item instrument. Forty-seven dog breeds were represented (Table 2). The result of the FA gave 4 factors with similar items loading on to the 4 factors that had been derived for the 46 item instrument during EFA (xxxx). The confirmatory factor solution accounted for 63% of the variance in the data which was similar to that of the

EFA (64%).

Reduction of items

Eight items – sluggish, confident, unsociable, contented, alert, obedient, reluctant, and frightened – were removed on the basis of their loading and communality in the confirmatory FA, and the remaining 38 items were considered for exclusion on the basis of their correlation with other domain items, their readability and their ability to discriminate between unwell and healthy dogs. Figure 1 shows the difference between an item that was judged independently by authors xx and xx to discriminate well between unwell and healthy dogs (A - uncomfortable) and one that they judged did not (B – subdued). Table 3 lists the 24 items removed from the 46-item long form instrument along with the reasons for their removal.

Factor analysis of the 22 items retained for the shortened instrument

The 4 factor model accounted for more of the variance than the 2 and 3 factor models and was the most interpretable model, with factors very similar in terms of their item loadings to those of the 46 item instrument. These 22 item factors were named as domains of HRQL by the authors in accordance with the items loading onto them (Energetic/Enthusiastic (E/E), Happy/Content (H/C), Active/Comfortable (A/C), Calm/Relaxed (C/R). The 4-factor model accounted for 72% of the variance. The scoring algorithm derived for these 4 domains of HRQL was based on item–factor associations. For ease of interpretation, all domains were named positively and the scoring algorithm provided that higher scores in all domains were associated with better

HRQL.

Field test of the shortened instrument

Owners of 53 unwell dogs and 100 healthy dogs (Table 1) completed 1 assessment and, of these, 49 owners of healthy dogs completed 2. Forty-two dog breeds were represented with no breed predominating (Table 2) A comparison of median scores and the interquartile range (IQR) for healthy and unwell dogs for each of the domains (Table 4) showed clear differences for E/E, H/C and A/C, but less so for C/R. However, the results of the Mann Whitney tests (Table 5) demonstrated a significant difference ($p=<0.05$) between the scores for healthy and unwell dogs in all domains. The variability, represented by the IQR and the extent of the tails of the distribution, was large in all domains for the unwell dogs compared with that of the healthy dogs, with the exception of C/R where the variability was similar between the groups (Figure 2). Linear discriminant analysis showed that the 22 item short instrument correctly classified 89% of the healthy dogs and 77% of those that were unwell with an overall misclassification rate of 15%. For those owners who completed 2 assessments the ICC (95% confidence intervals) for all domains was Energetic/Enthusiastic 0.75 (0.60 – 0.85); Happy/Content 0.75 (0.60 – 0.85); Active/Comfortable 234 0.75 (0.60 – 0.85); Calm/Relaxed (0.57 – 0.84).

Discussion

A review article by Goetz et al in 2013 concluded that item reduction of an existing scale must be based on rigorous methodology if the short-form instrument aims to maintain the validity and other measurement properties of the parent instrument. To that end they highlighted the importance of reporting the validity of the original scale, documenting the reasons for item selection, preserving content validity and the psychometric properties of the original scale and validating the short-form scale in an independent sample. The shortening process described here followed these guidelines.

Construct validity of the original 46 item long form instrument had been demonstrated using factorial validity and a known groups approach in dogs with a variety of chronic conditions, and evidence of its reliability had been obtained (xxxx). Additionally, these 46 items have been shown to be able to generate a valid measure of HRQL in dogs with osteoarthritis (unpublished) and lymphoma (unpublished), where both disease and aggressive treatment may impact on a dog’s QOL, and in obese dogs (Yam et al., 2016), all of which support the validity of the 46-item instrument. Furthermore, the validity of the original instrument was endorsed by the fact that the CFA performed as part of this study produced a factor structure in a new data set that was similar to the original in terms of factors, items and their loadings, and which accounted for a similar amount of the variance (63% vs 64%).

Factor loadings of 0.3, 0.5 and 0.7 are generally considered to be low, medium and high respectively (Shevlin & Miles, 1998) with loadings of >0.3 deemed to be the minimum consideration level for exploratory factor analysis. Exclusion of items with loadings <0.3 is commonly applied in instrument development (Floyd & Widaman, 1995). However, loadings of >0.5 are considered to be practically significant and accordingly the first step in the item reduction process described here was to exclude items which loaded < 0.5. It has been suggested that factor structures are improved when both loadings and communalities are higher (Velicer et al., 1982; MacCallum et al., 1999) and so items with a communality <0.5 were excluded as part of the shortening process. Thirty-three percent (8/24) of the removed items were excluded by this initial process and calculation of the Pearson Coefficient to exclude items that were too similar and therefore extraneous accounted for a further 29% (7/24). Once highly correlated items were identified, the process of choosing which to keep depended on their discrimination and readability. For example, ‘pained’ and ‘sore’ had a correlation of 0.84, both distinguished well between healthy and unwell dogs, but on readability grounds ‘sore’ performed better than ‘pained’, so ‘pained’ was excluded and ‘sore’ retained. The groups used to test readability in the USA and Australia were small (9 & 15 dog owners respectively) and were not representative of the general dog owning population which could be seen as a weakness in the study, but all 46 words had been pretested previously in the UK (xxxx) and the purpose of the Australian and USA tests was purely to identify cultural difference in relation to meaning; accordingly, the groups were considered adequate.

For the purpose of establishing known groups construct validity, expert judgement has been used previously to identify items that could discriminate between unwell and healthy dogs (xxxx) and that process was repeated here. Although it was considered unlikely that an item that was unable to discriminate well from unwell dogs would prove useful in an evaluative context, that possibility cannot be discounted and removed items may be reassessed if the instrument proves not to be responsive to clinical change in further longitudinal studies, in order to develop a longer-form instrument for evaluative purposes.

Factor analysis to determine the optimum factor model on which to base the HRQL domains of the shortened 22 item instrument was carried out on results from 122 dogs (88 unwell and 34 well). The literature includes a range of recommendations regarding the minimum sample size necessary to obtain factor solutions that are adequately stable, including the suggestion that there should be between 4 and 5 times as many samples as variables (Floyd & Wideman, 1995). On that basis the sample size used here was adequate. However, several workers including Velicer and Fava (1998), found the influence of sample size to be reduced when factor loadings and communalities were high, which was the case in our study where all loadings were >0.5 and 14/25 were >0.7 which is considered high. Similarly, 12/22 communalities were >0.7 and according to MacCallum et al (1999) communalities of 0.6 are considered high. Consequently, the factor model was considered stable.

A useful factor model captures a reasonable amount of the total variance in the data from which it is derived, with higher figures representing better models. A perfect model would account for 100% of the variance in the sample, but this would have the same number of factors as variables and Norman & Streiner (1994) suggested that factors should explain at least 50% of the total variance. The 64% and 62% of the variance captured by EFA and CFA of the 46-item instrument compare well with that accounted for in other proxy instruments to measure the QOL of infants (45%) (Manificat et al., 1999), the QOL of older children (62%) (Varni et al., 2001), the behaviour and temperament of guide dogs (63%) (Serpell & Hsu, 2001) and of pet dogs (57%) (Hsu & Serpell, 2003). However, the 22 items comprising the shortened instrument accounted for 72% of the variance compared with 62% for the CFA of the 46 items using the same data set, indicating an improvement in the factor model. This could be a result of the higher loadings (>0.5 vs >0.3) representing a closer association of the 22 items with the factors compared with that of the 46 items as a result of the removal of less correlated items which had contributed some measurement 'noise'. Further to this demonstration of factorial validity, field-testing of the new instrument was designed to confirm that shortening of the instrument had not diminished the psychometric properties of the original. Known groups validity was demonstrated by the fact that scores in all 4 domains of HRQL were significantly different between healthy and unwell dogs. In common with the 46 item instrument, the domain scores in the unwell dog group showed more variation than those for the healthy dogs. The study protocol did not ask clinicians to rate the severity of disease in the unwell dogs, but only specified that cases should be selected on the basis that the condition was likely to affect the

QOL. However the wide interquartile range and the extent of the whiskers in Fig 2 in the unwell dogs would tend to suggest that there was a wide spread of disease severity with resultant variability in their health status. Subjective evaluation of general behavioural signs such as changes in appetite, activity and sociability have long been reported as changing with ill-health, especially in food animals (Weary et al., 2009), but to the authors' knowledge the HRQL domains reported here - Energetic/Enthusiastic, Happy/Content, Active/Comfortable and Calm/Relaxed - have not been specifically reported in companion animals as likely to change with health status. However the SF 36 is a generic HRQL instrument for people designed to measure physical and emotional components of health status and it contains the terms 'activity', 'calm and peaceful', 'full of life', 'energetic' and 'happy' (Ware, 1992). The domain Calm/Relaxed shows more variability in the healthy dogs than was apparent for the other 3 domains, which is perhaps not surprising given the spectrum of excitability in the healthy dog population. There is also a smaller difference in median scores between the healthy and unwell groups in that domain and more overlap in the interquartile ranges. This may be accounted for by the fact that this domain contains items (eg 'calm') that could reflect relatively stable personality traits, making it more resistant than other domains to change with ill health.

Owners of healthy and unwell dogs for CFA, item reduction and EFA of the 22 item shortenedinstrument and the field test for the 22 item instrument were drawn in part from a university referral population which may raise some concerns regarding respondent bias. However the authors consider that drawing from a variety of sources where possible (1y care and referral) broadens the scope of the recruitment in clinical studies where it is very difficult to control or selection bias. With respect to the use of vets and vet nurses as respondents, who might be influenced by their professional expertise, the questions in the instrument are related to owner observed behaviours and do not involve any judgement related to health or welfare. Also the university staff who took part in the study included a mix of administrative and non–veterinary teaching staff in addition to vets and nurses. Because the answers to the questions in the instrument involve an interpretation of behaviour on the part of the owner and that interpretation is best made by the person that knows the dog best, only the primary carer of the dog was recruited to the study. Additionally, where 2 assessments were carried out the same owner completed these.

Although there was a discrepancy between the types of cases included in each field test, one of which provided the data for the CFA, item reduction and EFA of the shortened instrument (primarily referral), with the other providing the data for the field test of the shortened instrument (primarily 1st opinion) all dogs were suffering from a condition, usually chronic, likely to affect their QOL. General practitioners regularly treat cases such as osteoarthritis, obesity, diabetes, cardiac failure, chronic skin disease and cancer in 1st opinion practice and accordingly the authors suggest that in this context the impact of any differences is not likely to be significant. The discriminant analysis with cross-validation of the 22 item short form indicated an overall misclassification rate of 15% with 89% of healthy dogs and 77% of unwell dogs classified correctly. These results compare well with those reported for a proxy instrument for pain measurement in communicatively impaired children that correctly classified 92.9% of children with no pain and 71.3% of children in pain with an overall misclassification rate of 13% and which was considered by its developers to have reasonable ability to distinguish between pain and no-pain episodes (Stallard et al., 2002). Misclassifications in the study reported here may have been a result of measurement error, or may have occurred because the QOL of some of the healthy dogs was compromised at the time of measurement for reasons other than poor health, or because some of the dogs that were unwell may in any case have been experiencing a good quality of life at the time of measurement. For example an unpublished study demonstrated that a group of dogs diagnosed with multicentric lymphoma subtype A had, at the time of first presentation to xxxxxxxxxx significantly higher HRQL scores than a group with subtype B. It seems likely that all of these reasons for misclassification would be true to some extent. Scores on the instrument are not intended to replace clinical evaluation, but should be regarded as a valuable adjunct, replacing subjective owner report with valid and reliable measurement of change at one time point and over time.

The ICC values for the domains were >0–7, 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. This result indicated that the reliability of the shortened instrument was improved compared with the original where ICC values were >0.6. However, the current test was carried out with 49 owners compared with 16 for the original and this may have contributed to the improved result (xxxx)

Best practice in instrument design dictates that when a questionnaire instrument is presented in a new way, for example moving from paper to web, or re-design of presentation, it is recommended that the instrument be re-tested in its new form to ensure that changes in format or design have not altered its measurement properties. For logistical reasons, field testing of the shortened 22 item instrument was carried out using a web-based platform and not via a mobile phone application. However, since this paper was submitted for publication initially the shortened instrument has been incorporated in a smartphone app and has been shown to discriminate well between healthy and unwell dogs (unpublished).

In conclusion, the measurement of companion animal HRQL has much to offer the veterinary practitioner in terms of improved client communication and relations. This study has provided evidence for the instrument's ability to detect the HRQL impact of disease and work is ongoing to establish its usefulness in therapeutic monitoring and as a tool to facilitate the identification of humane endpoints for individual dogs. These capabilities are becoming more and more necessary in clinical practice as medical advances facilitate the keeping of animals with painful chronic disease for longer, and evidence based medicine requires that robust measures of clinical impact be developed. This study has provided initial evidence for the reliability and validity of the shortened instrument. However, it is important to emphasise that validity is not determined by a single statistic, but by a body of research supporting claims that the instrument is valid for particular purposes, with defined populations and in specified contexts (Streiner and Norman, 2008). Accordingly, future research will seek to provide such evidence, including evidence for its responsiveness in longitudinal studies.

References

Abell, N., Springer, D. W., Kamata, A. (2009) Developing and Validating Rapid Assessment Instruments. New York: Oxford University Press. pp 10

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

Boulos, M. N. K., Brewer, A. C., Karimkhani, C., Buller, D. B., & Dellavalle, R. P. (2014). Mobile medical and health apps: state of the art, concerns, regulatory control and certification. Online Journal of Public Health Informatics, 5(3), 229

Boyle, G. J. (1991). Does item homogeneity indicate internal consistency or item redundancy in psychometric scales? Personality and Individual Differences, 12(3), 291–294.

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

Brown D.C. (2014) The Canine Orthopaedic Index. Step 1: Devising the Items.

Veterinary Surgery 43, 232-224

Brown D.C (2014) The Canine Orthopaedic Index. Step 2: Psychometric Testing. Veterinary Surgery 43, 241-227

Brown D.C (2014) The Canine Orthopaedic Index. Step 3: Responsiveness Testing. Veterinary Surgery 43, 247-254

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

Cheak-Zamora, N.C., Wyrwich, K.W. and McBride, T.D., 2009. Reliability and validity of the SF-12v2 in the medical expenditure panel survey. Quality of Life Research, 18(6), pp.727-735.

Coste J, Walter E, Venot A. (1995) A new approach to selection and weighting of items in evaluative composite measurement scales. Stat Med; 14: 2565-2580.

Coste, J., Guillemin, F., Pouchot, J., & Fermanian, J. (1997). Methodological approaches to

shortening composite measurement scales. Journal of Clinical Epidemiology, 50(3), 247–252.

Coste J, Bouée S, Ecosse E, et al. (2005) Methodological issues in determining the dimensionality of composite health measures using principal component analysis: case illustration and suggestions for practice. Qual Life Res. 14:641–654.

DeVellis, R. F. (2011). Factor Analysis. Scale Development: Theory and Applications, 45(2), 115–158

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

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

Fayers, P.M. and Machin, D., 2013. *Quality of life: the assessment, analysis and interpretation of patient-reported outcomes*. John Wiley & Sons.

Floyd, F. J., & Widaman, K. F. (1995). Factor Analysis in the Development and Refinement of Clinical Assessment Instruments. *Psychological Assessment1*, *7*(3), 286–299.

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 the American Veterinary Medical Association 226, 1864-1868

Goetz, C., Coste, J., Lemetayer, F., Rat, A. C., Montel, S., Recchia, S., - Guillemin, F. (2013). Item reduction based on rigorous methodological guidelines is necessary to maintain validity when shortening composite measurement scales. Journal of Clinical Epidemiology, 66(7), 710–718.

Greenwood, M. C., Hakim, A. J., Carson, E., & Doyle, D. V. (2006). Touch-screen computer systems in the rheumatology clinic offer a reliable and user-friendly means of collecting quality-of-life and outcome data from patients with rheumatoid arthritis. Rheumatology, 45(1), 66–71.

Hsu Y, Serpell JA. Development and validation of a questionnaire for measuring behavior and temperament traits in pet dogs. J Am Vet Med Assoc 2003;223:1293–1300

Johnston CC. Psychometric issues in the measurement of pain. In: Finley GA, McGrath PJ, eds. Measurement of pain in infants and children: progress in pain research and management. Vol 10. Seattle: IASP Press, 1998;5–20

Krebs, P., & Duncan, D. T. (2015). Health App Use Among US Mobile Phone Owners: A National Survey. JMIR mHealth and uHealth, 3(4), e101.

Kane, L. R. (2006) 2nd edn. Eds L. R. Kane, Jones and Bartlet Learning, Massachusetts

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, 1-12

Lavan, R.P., 2013. Development and validation of a survey for quality of life assessment by owners of healthy dogs. The Veterinary Journal, 197(3), pp.578-582.

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 -18

Manificat S, Dazord A, Langue J, et al. A new instrument to evaluate infant quality of life. MAPI Res Inst Qual Life Newslett 1999;23:7–8.

MacCallum, R. C., Widaman, K. F., Zhang, S. B., & Hong, S. H. (1999). Sample Size in Factor Analysis. *Psychological Methods*, *4*(1), 84–99.

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

Norman, G.R. and Streiner, D.L., 1994. Principal components and factor analysis. Biostatistics, the Bare Essentials. St. Louis, MO, Mosby-Year Book, pp.129-142.

Nunnally J.C., & Bernstein I,H. (1994) Psychometric Theory (3rd ed.). New York: McGraw-Hill,

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

Reid, J., Wiseman-Orr, M. L., Scott, E. M.,et al. (2013) Development, validation and reliability of a web-based questionnaire to measure health-related quality of life in dogs. Journal of Small Animal Practice 54, 227 - 233

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

Serpell JA, Hsu Y. Development and validation of a novel method for evaluating behavior and temperament in guide dogs. Appl Anim Behav Sci 2001;72:347–364

Shevlin, M., & Miles, J. N. V. (1998). Effects of sample size, model specification and factor loadings on the GFI in confirmatory factor analysis. *Personality and Individual Differences*, *25*, 85–90.

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

Stallard, P., Williams, L., Velleman, R., Lenton, S., McGrath, P. J., & Taylor, G. (2002).

The development and evaluation of the pain indicator for communicately impaired children (PICIC). Pain, 2, 149.

Stanton, J. M., Sinar, E. F., Balzer, W. K., & Smith, P. C. (2002). Issues and strategies for reducing the length of self-report scales. *Personnel Psychology*, *55*(1), 167–194.

StatSoft Inc. Principal components and factor analysis 1984–2003). Available at: http://www.statsoft.com/textbook/stfacan.html . Accessed April 3rd 2017

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. pp 399

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

Varni JW, Seid M, Kurtin PS. PedsQL 4.0: reliability and validity of the Pediatric Quality of Life Inventory version 4.0 gener- ic core scales in healthy and patient populations. Med Care 2001;39:800–812.

Velicer, W. E, Peacock, A. C., & Jackson, D. N. (1982). A comparison of component and factor patterns: A Monte Carlo approach. Multi- variate Behavioral Research, 17, 371-388.

Velicer, W. F., & Fava, J. L. (1998). An Evaluation of the Effects of Variable Sampling On Component, Image, and Factor Analysis. Psychological Methods, 3(2), 231–251.

Ware J.E (1992) The MOS 36-Item Short-Form Health Survey (SF-36): I Conceptual Framework and Item Selection Author (s): John E . Ware , Jr . and Cathy Donald Sherbourne Published by Lippincott Williams & Wilkins Stable Medical Care, 30 (6) 473

– 483

Weary D.M, Huzzey J. M. & von Keyserlingk M. A. G. (2009) BOARD-INVITED REVIEW: Using behavior to predict and identify ill health in animals. J Anim Sci 2009.87:770-777.

Wiseman ML, Nolan AM, Reid J and Scott EM (2001) Preliminary study on owner-reported behaviour changes associated with chronic pain in dogs. Veterinary Record *149*: 423-524

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

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

Wojciechowska, J. I., Hewson, C. J., Stryhn, H., Guy, N. C., Patronek, G. J., & Timmons, V (2005). Development of a discriminative questionnaire to assess nonphysical aspects of quality of life of dogs. *American Journal of Veterinary Research*, *66*(8), 1453–1460.

Wojciechowska JI, Hewson CJ, Stryhn H, et al. Evaluation of a questionnaire regarding nonphysical aspects of quality life in sick and healthy dogs. Am J Vet Res 2005;66:1460–1467.

Yam, P. S., Butowski, C. F., Chitty, J. L., Naughton, G., Wiseman-Orr, M. L., Parkin, T., & Reid, J.(2016). Impact of canine overweight and obesity on health-related quality of life. Preventive Veterinary 508 Medicine, 127, 64–69

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