Development, initial validation and reliability testing of a web-based, generic feline health-related quality of life instrument

Dr Noble, CE - *NewMetrica Ltd, 19 Woodside Place, Glasgow, G3 7QL, UK*

Dr Wiseman-Orr, ML- *School of Mathematics and Statistics, 15 University Gardens, University of Glasgow G12 8QW, UK*

Prof Scott, EM- *School of Mathematics and Statistics, 15 University Gardens, University of Glasgow G12 8QW, UK*

Prof Nolan, AM- *Edinburgh Napier University, Sighthill Campus, Sighthill Court EH 11 4BN, UK*

Prof Reid, J -*NewMetrica Ltd, 19 Woodside Place, Glasgow, G3 7QL, UK; College of Medical, Veterinary and Life Sciences, University of Glasgow G12 8QQ, UK.*

Corresponding author:
Prof Reid, BVMS PhD DVA DipECVAA MRCA MRCVS

jacky.reid@newmetrica.com

University of Glasgow
CIDS Incubator, Rm 203
Thomson Building
University Avenue
Glasgow, G12 8QQ

+44 (0) 7876683262

Abstract

Objectives: To develop a valid, reliable web-based, generic feline health-related quality of life (HRQL) questionnaire instrument to measure the affective impact of chronic disease.

Methods: A large initial item pool, obtained through interviews with cat owners, was reduced using pre-determined criteria, survey scores for relevance and clarity and the ability of individual items to discriminate between healthy and sick cats when owners completed a prototype questionnaire. Using these data, factor analysis was used to derive a scoring algorithm and provide evidence for factorial validity. Validity was demonstrated further in a field trial using a "known groups" approach (sick vs healthy cats will have a different HRQL profile, and the HRQL profile of cats will deteriorate as co-morbidities increase). Test–retest reliability was assessed using intraclass correlation coefficients (ICC).

Results: One hundred and sixty five items were reduced to 20 and, on the basis of a factor analysis that explained 72.3% of the variation in scores input by 71 owners of 30 healthy and 41 sick cats using the prototype, these were allocated to three domains [vitality, comfort and emotional wellbeing (EWB)] with a scoring algorithm derived using item loadings. Subsequently, owners of 36 healthy and 58 sick cats completed 1 or 2 (48) assessments. Median scores (healthy vs sick) for all domains were significantly different ($p=<0.001$), 78% of cats were correctly classified as healthy or sick and for comorbidities the correlation coefficients were moderate (vitality – 0.64; comfort – 0.63; EWB - 0.50). Test-retest reliability was good (ICC – vitality 0.635; comfort 0.716; EWB 0.853)

Conclusions and relevance: This study provides initial evidence for the validity and reliability of a novel HRQL instrument to aid the assessment and management of chronic diseases of cats.

Introduction

It is the unpleasant feelings (affective component) associated with chronic disease that cause an individual to suffer. The medical profession recognises the importance of the valid and reliable measurement of how people feel and has addressed this through the development of instruments to measure health-related quality of life (HRQL) for disease detection (discriminative purposes) or to measure change in health status over time (evaluative purposes). (1) Structured questionnaire instruments are developed and tested using well established psychometric methodology. (2–4)

Instruments to measure HRQL in companion animals consist of questions for the owner, who is well placed to report upon the subtle changes in behaviour, attitude and demeanour that occur with chronic disease. While several feline disease-specific instruments exist (5–9), to date, no validated generic HRQL instrument exists for the purpose of comparing treatments or disease states. (10) A generic instrument (CHEW), based on owner perceived health status has been reported (11), but not validated in sick cats. Similarly CatQol (Bijsmans) focuses on general health, eating, behavior and management, but has been validated in cats with chronic kidney disease (CKD) only. (12) More recently Tatlock et al (2017) have described an owner reported feline quality of life scale for healthy cats. (13)

Briefly, the psychometric approach to instrument design consists of generation of a pool of items (questions) most often though interviews with key informants, reduction of these by various techniques including expert judgement of the relevance and adequacy of items (14), the identification of items that do not discriminate well between known groups of subjects (15) and the use of a statistical technique called Factor Analysis (16), before pretesting and then testing for validity and reliability.

Evidence for any new instrument's validity and reliability is essential before use in a clinical context. Various kinds of validity may be sought. For example, content validity is a measure of the extent to which the items included in a questionnaire are relevant and adequate for its purpose: it is established during its construction and assessed by expert judgement. Criterion validity is the agreement of a new instrument with some existing 'gold standard', but where that does not exist, evidence can be gathered to support concurrent criterion validity (comparison with a validated measure of a related construct) or predictive criterion validity where performance of the new measure successfully predicts that of a later measure. Construct validity – evidence that the instrument is measuring the construct that it is intended to measure - is considered to be the most robust and fundamental form of validity. (1) A construct is something that is not directly observable or measurable, such as "happiness". The construct being measured here is HRQL which is the subjective evaluation by an individual of its circumstances that include an altered health state and the impact of related interventions. (17) Construct validity is established by a process of hypothesis testing, where hypotheses are based upon how an instrument should perform if it is measuring the construct of interest. For example, factorial validity applies if factor analysis of data generated using the instrument reveals an interpretable factor structure that fits the construct the instrument was designed to measure. (18) In a 'known-groups' approach to construct validation, predictions are made about how scores obtained with the instrument will differ between groups, such as healthy and sick animals, or will reflect disease burden, and these predictions are tested.

A reliable instrument 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). (2)

Previously a novel generic instrument to measure HRQL in dogs was developed in which most of the items reported aspects of behaviour that owners believed were expressions of a dog's subjective experience (feelings) (17,19), and evidence for the validity and reliability of a web-based version was reported. (15) Subsequent shortening resulted in a 22-item instrument that retains the capacity of the prototype to measure the animal's feelings. (20) The aim of this study was to develop an equivalent generic instrument for cats and to provide first evidence for its content validity, construct validity and intra-rater reliability.

Materials and methods

Ethical approval was granted by the University of Glasgow, and all participants gave informed consent.

Item generation and initial selection

Semi-structured interviews were conducted with owners of healthy cats and cats with conditions likely to affect quality of life (QOL), recruited through the University of Glasgow Small Animal Hospital (UGSAH) and several veterinary practices. The interviews were recorded and transcribed verbatim to generate items consisting of terms and phrases used by interviewees to describe their cats when healthy and sick. Interviews were continued until no new information emerged and redundancy was reached. (2,20) Questions (open and closed) were worded carefully to limit response bias. (21,22) Qualitative analysis of the transcripts was conducted using established methods in grounded theory, a methodology commonly used in the social sciences which involves the gathering and analysis of data to construct a theoretical framework for whatever is being studied. This is in contrast to

conventional methods which adopt an existing theoretical framework, and then collect data to show whether or not it applies to the phenomenon being studied. (23)

Each item was considered by the authors and excluded if it was deemed to be related to individual personality traits; disease specific; lacking in clarity/readability; more relevant to clinical examination rather than owner report; not relevant to HRQL; or where a more appropriate description of that observation had been offered.

Content validation

An online survey (SurveyMonkey Inc., San Mateo, California, USA) of the remaining items was conducted in which groups of cat owners and clinicians were asked to judge their clarity and their relevance to the measurement of feline HRQL. Relevance was scored using a four-point Likert scale (0 - "not relevant", 3 - "very relevant"). (24) The clarity of each item was determined using binary response options, "clear or not clear".(25) Participants were asked for feedback on why they found an item not relevant or unclear, and were invited to suggest additional items.

Degree of relevance scores were then dichotomized by assigning scores of 0 and 1 as "not relevant" and scores of 2 and 3 as "relevant". Content validity Index scores for relevance (I-CVI_R) and clarity (I-CVI_C) of each item were derived by averaging the scores given to each item and dividing by the number of respondents(24). Items were excluded if I-$CVI_R \leq 0.60$ or considered for revision/exclusion if I-$CVI_C < 0.70$.

Prototype construction and pre-testing

In the prototype instrument, each item was accompanied by a 7-point Likert scale, 0 to 6 (0, "not at all"; 6, "couldn't be more") to allow the respondent to rate the extent to which the item described his or her cat. Some of these anchors were re-worded to suit the form (word or phrase) of the item.

Software developers, Kyria Ltd (http://Kyria.co.uk) produced a web-based prototype instrument for pre-testing with a number of cat owners. The prototype was revised as required to ensure optimal utility, functionality and lack of ambiguity.

To compare owners' impression of health status compared with that of clinicians, which has been found to differ in the dog according to previous work (unpublished) an additional owner question "is your cat perfectly healthy?" – Yes/No was included in the prototype instrument, but did not form part of the assessment.

Field Test 1 for item reduction, factorial validation and determination of scoring algorithm

Cat owners were recruited from first opinion practices, feline specialist practices and the University of Glasgow Small Animal Hospital (UGSAH). Owners completed one assessment for their cat and the attending clinician completed a general health assessment (Appendix 1) to verify the cat's health status.

The research team reviewed dot plots of the response scores (0 to 6) generated for each of the items and eliminated any item that was judged by all not to discriminate well between healthy and sick cats.

To establish evidence for factorial validity, and to determine a scoring algorithm for the instrument, a factor analysis (principal components method with a varimax rotation) was performed with remaining items. The scores attributed to each item by the owners were used for the factor analysis. The analysis allocates each item to a factor with a loading (0 – 1) which determines the closeness of its relationship to the factor. Resulting loadings were sorted, and any item with a loading <0.4 was excluded. A scree test and the Kaiser criterion were used to identify the optimum number of factors and 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 a factor model was chosen 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. 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 2 for construct validity and reliability testing

A new group of cat owners was recruited from first opinion practices, feline specialist practices and the University of Glasgow Small Animal Hospital (UGSAH). The attending clinician completed a general health assessment (Appendix 1) to verify the cat's health status.

Owners of healthy and sick cats, grouped according to the clinical judgement of the consulting clinician, completed at least one assessment. A number of owners of healthy cats 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 (cats) are assumed random. (26)

Using the first assessment for each cat, box plots and descriptive statistics were used to identify differences between healthy and sick cats, followed by formal statistical analysis using non-parametric Mann-Whitney tests due to the non-normality of the data. Linear discriminant analysis was used to determine the ability of the instrument to differentiate healthy from sick cats. The correlation between the number of comorbidities affecting each cat and their HRQL scores was investigated using linear regression and Pearson Correlation Coefficient for all 3 domains was calculated for healthy cats, cats with 1-2 comorbidities and cats with ≥3 comorbidities.

The following hypotheses were tested: a) that the HRQL profile of scores will differ between healthy cats and sick cats and b) that the HRQL profile will be worse for cats with poorer health status as defined by the number of comorbidities present in individuals.

Results

An overview of the development process is shown in Figure 1

Item generation and initial selection

Semi-structured interviews conducted with the owners (n=18) of healthy (n=19) and sick (n=10) cats (Table 1) generated an initial pool of 165 items for consideration. Table 2 illustrates the format of interview questions. One hundred and six items met the authors' criteria for exclusion or revision (for examples see Table 3). Fifty-one items were retained for content validation.

Content validation

Fifty-eight participants (48 owners of 14 sick cats and 34 healthy cats) and 10 clinicians completed surveys assessing the clarity and relevance of the remaining 51 items. As a result of not having met I-CVI criteria, 13 of these items were eliminated and 11 items were revised, with one of those being split into two separate items (Figure 1)*.*

Prototype construction and pre-testing

The prototype for Field Test 1 consisted of 39 items, 27 single words with the standard response option (0 -"not at all"; 6 - "couldn't be more") and 12 items where response options were reworded to suit the form of the item (e.g. "hiding away" – Not hiding away at all/ Couldn't be hiding away more).

Pre-testing of the online instrument was conducted with 15 owners of 5 healthy cats and 10 sick cats. Following this, response options for 2 items – 'Jumping or climbing up/down', and "Usual sleeping patterns and/or places" were revised to improve readability and comprehension.

Field Test 1 for item reduction, factorial validation and determination of scoring algorithm

Using the online prototype instrument for Field Test 1, 71 owner and clinician assessments from UGSAH, 5 general practices and 1 feline specialist practice were completed over a period of 5 months for 30 healthy cats and 41 cats diagnosed with a chronic condition expected to impact QOL (Table 1). Ninety-five percent of cats presented with 1 to 6 comorbidities (Table 5). Review of dot plots of these item responses suggested that 19 items were unlikely, in the opinion of the research team, to discriminate between sick and healthy

220 cats (Figure 2); these were removed leaving 20 items to be included in the instrument for
221 Field Test 2 (Table 4).
222 A factor analysis was conducted using the responses to these 20 items, all of which had
223 loadings >0.4. A model containing three factors was considered to be optimal, accounting
224 for 72.3% of the variance in the owner response data and consisting of factors that could be
225 interpreted as HRQL domains which were named by the research team as "vitality" (11
226 items), "comfort" (8 items) and "emotional wellbeing" (EWB) (7 items). Some items loaded
227 onto more than one factor. An algorithm, based on the item–factor associations for the 3-
228 factor model, was derived in order to generate three domain scores. However, for
229 commercial reasons a description of the factor composition and the algorithm are not
230 presented.

231 ***Field Test 2 for construct validity and reliability testing***

232 Using the resulting online instrument for Field Test 2, the owners of 36 healthy cats and 58
233 sick cats as determined by clinician assessment, representing a comprehensive range of
234 breeds (Table 1), completed 1 assessment and, of these 94 owners, 48 owners completed 2
235 assessments. According to their responses to a direct question, a total of 26% of owners of
236 sick cats believed their cat to be perfectly healthy, despite a clinician diagnosis of ill health.
237 Sick cats presented with a range of conditions and 72% had 1 to 6 additional comorbidities
238 (Table 5).
239

240 **Construct validity**

241 Differences between healthy and sick cats existed for all three domains (Table 6) supporting
242 hypothesis 1 (null hypothesis: no difference in median score between healthy and sick cats,
243 rejected at p-value <0.01), with greater variability in the sick group compared with the

healthy group (Figure 3). Linear discriminant analysis (using cross validation) showed that the instrument correctly classified as either healthy or sick 78% (healthy, 71%; sick 89% classified correctly) of the cats assessed. An increase in number of comorbidities was associated with a deterioration in HRQL profile (Figure 4). The Pearson Correlation Coefficients for Vitality, Comfort and Emotional Wellbeing and comorbidities (healthy, 1-2 and >3) were -0.64, -0.63 and -0.50 respectively.

Reliability

Forty owners completed a second assessment for their cats with a minimum of 14 days between assessments and the ICCs (95% confidence intervals) for Vitality, Comfort and Emotional Wellbeing were vitality 0.635 (0.044 to 0.862), comfort 0.716 (0.256 to 0.893) and EWB 0.853 (0.615 to 0.945).

Discussion

One hundred and sixty-five potential items were collected from the owners of sick and healthy cats using best practice for qualitative research. (2) Since information obtained from key informants underpins content validity, comprehensive representation of all relevant populations is necessary. Freeman et al (2016) described a generic HRQL scale for cats (11), where key informants were restricted to owners/caregivers of healthy cats. Similarly Tatlock et al (2017) used pet owners of healthy cats as key informants. (13) However 29% and 27% of owners (Field Tests 1 and 2) in this study thought their cats were healthy when clinicians deemed them to be sick, reinforcing that such judgement may be unreliable. (27) Bijsmans et al (2016) used owners as informants for CatQol, but no details are available regarding the health status of these cats. In contrast, the health status of the

19 healthy cats and 10 sick cats belonging to owners recruited as key informants in this study was verified by a veterinary surgeon. Although this number of cats could be considered to be low, interviews were conducted until no new information emerged. In addition, 48 different owners involved in the content validation process were invited to suggest additional items if they felt the collection of items was inadequate. Initial reduction of the 165 items was based on criteria devised by the investigators, an approach considered to be appropriate in human medicine. (28–30)

In veterinary science, many rely on cognitive debriefing interviews to establish the content validity of an instrument scale (11,31) or simply ask owners to judge whether the instrument appears to be capable of measuring what it is intended to measure (face validity). (32) However, in this study a group of vets as well as a large group of owners were involved in the validation process, adding to the robustness of this stage in the process.

In human medicine and the social sciences, the quantification of content validity has been introduced. One approach, used here, asks relevant 'experts' to rate the relevance and clarity of items using a rating scale, and those ratings are used to calculate a CVI for each individual item on the scale (I-CVI), providing objective information to guide researchers in revising, deleting, or substituting items. (33) The instrument described here is the first instrument in veterinary science to quantify and establish content validity of each item using this technique.

Following Field Test 1, 19 items were excluded based on research team judgement that they did not discriminate healthy from sick cats. Although it was considered unlikely that an item that was unable to discriminate healthy from sick cats would prove useful in an evaluative context, that possibility cannot be discounted and items removed at this stage may be

reassessed for inclusion if the instrument proves not to be responsive to clinical change in further longitudinal studies.

The remaining 20 items all loaded >0.4 in the FA. Factor loadings of 0.3, 0.5 and 0.7 are generally considered to be low, medium and high respectively (34) with loadings of 0.3 deemed to be the minimum consideration level for FA. Increasing the loading threshold to 0.4 may have accounted for the fact that no further items were removed at this stage. Furthermore, the fact that the majority of loadings were >0.6 indicates stability of the factor model. (35) Although FA provides any number of factor models for a given data set, there are established methods of identifying how many factors could sensibly be extracted, including the scree test and the Kaiser criterion (36), both considered in this study. The 3 factor model adopted here, accounting for 72% of the variability in the owner response data, compares favorably with the canine HRQL (64%)(15) and 72% for the shortened instrument (20), an 11-factor questionnaire designed to measure the behaviour and temperament of pet dogs (57%) (37) and a 4-factor QoL questionnaire regarding infants (38), and confirms factorial validity.

Scores in all 3 domains of HRQL were significantly different between healthy and sick cats. The fact that domain scores in the sick cats showed more variation than the healthy cats is not surprising given the heterogeneity of disease in the sick cat population. Interestingly, the variability in healthy cat vitality was similar to that of the sick cats, but this probably reflects the tendency for the individual variation in vitality which tends to exist in this species. Furthermore, cats with greater comorbidity had lower HRQL scores indicating a poorer QOL, with moderate Pearson Correlation Coefficients for all 3 domains (39), thus

upholding known groups hypothesis (b), providing additional evidence for construct validity. Evidence for known group validity relating to health status of other generic HRQL instruments in cats is sparse. Freeman et al (2016) (11) investigated the validity of their scale in a large group of 1303 cats but only 8 of these were categorised as being 'not very healthy' or 'not at all healthy' by their owners, in any case a judgement that we have shown in this study to be unreliable. Bijsmans et al., (2016) (12) demonstrated that their instrument detected difference between healthy cats and those with CKD, but this evidence is of limited value in relation to the proposed generic nature of their instrument. Discriminant analysis indicated an overall misclassification rate of 22% compared with that reported for dogs with chronic pain (12% misclassification) (18) and for a proxy instrument for pain measurement in communicatively impaired children (13%). (40) Misclassifications in the study reported here may have been a result of measurement error, or may have occurred because the QOL of some healthy cats was compromised at the time for reasons other than poor health, or because some sick cats may have been experiencing a good QOL at the time.

Criterion validity was not carried out because no gold standard instrument for the measurement of HRQL exists, but the authors do not discount the possibility of being able to demonstrate concurrent or predictive criterion validity in future studies when suitable measures become available.

Test-retest reliability was carried out on data for healthy cats only, whose health status would be less likely to change over a 2-week period than would that of sick cats. A 2-week period between the completion of questionnaires is commonly chosen for this purpose, being a short enough period for change in health status to be unlikely while being a long enough period for respondents to be unlikely to remember their previous responses. The

ICC values for the Comfort and Emotional Wellbeing domains were >0.7 and >0.8 respectively, indicating that test–retest reliability for those domains was good, and it was moderate for Vitality (ICC >0.6). (41)

In conclusion, the measurement of feline HRQL is becoming more necessary as chronic diseases such as CKD, hyperthyroidism, cognitive decline and OA affect the QOL of an increasing number of ageing cats, 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 a novel generic instrument that measures the affective component of the chronic disease experience. However it is important to emphasise that validity is not determined by a single statistic, but by a body of research that supports the claim that the instrument is valid for particular purposes, with defined populations and in specified contexts. (2) Accordingly, future research will seek to provide such evidence, as well as evidence for the instrument's responsiveness to clinical change including that following treatment. The instrument is available for clinical use and for clinical trials from NewMetrica (www.newmetrica.com). For further information please contact the corresponding author, jacky.reid@newmetrica.com.

Acknowledgments

The authors wish to thank all the cat owners as well as the veterinary surgeons and nurses in practice and at the University of Glasgow Small Animal Hospital who willingly participated in our studies

Funding

Scottish Enterprise for the award of a SMART grant which enabled us to undertake the study, and to Boehringer Ingelheim for additional financial support.

Conflict of Interests

Professor Reid is a shareholder of NewMetrica Ltd, which is the developer and supplier of the instrument

References

1. Fayers PM, Machin D. *Quality of life : the assessment, analysis and interpretation of patient-reported outcomes*. John Wiley & Sons, 2013.
2. Streiner DL, Norman GR, Cairney J. *Health Measurement Scales*: *a practical guide to their development and use.* Oxford University Press, USA, 2015.
3. Abell N, Springer DW, Kamata A. *Developing and Validating Rapid Assessment Instruments*. Oxford University Press, 2009.
4. Brod M, Tesler LE, Christensen TL. Qualitative research and content validity: developing best practices based on science and experience. *Qual Life Res* 2009; 18: 1263–1278.
5. Freeman LM, Rush JE, Oyama MA, et al. Development and evaluation of a questionnaire for assessment of health-related quality of life in cats with cardiac disease. *J Am Vet Med Assoc* 2012; 240: 1188–1193.
6. Zamprogno H, Hansen BD, Bondell HD, et al. Item generation and design testing of a questionnaire to assess degenerative joint disease–associated pain in cats. *Am J Vet Res* 2010; 71: 1417–1424.
7. Niessen SJM, Powney S, Guitian J, et al. Evaluation of a Quality-of-Life Tool for Cats with Diabetes Mellitus. *J Vet Intern Med* 2010; 24: 1098–1105.
8. Noli C, Minafò G, Galzerano M. Quality of life of dogs with skin diseases and their owners. Part 1: development and validation of a questionnaire. *Vet Dermatol* 2011; 22: 335–343.
9. Noli C, Borio S, Varina A, et al. Development and validation of a questionnaire to evaluate the Quality of Life of cats with skin disease and their owners, and its use in 185 cats with skin disease. *Vet Dermatol* 2016; 27: 247-e58.
10. Giuffrida MA, Kerrigan SM. Quality of life measurement in prospective studies of cancer treatments in dogs and cats. *J Vet Intern Med* 2014; 28: 1824–9.
11. Freeman LM, Rodenberg C, Narayanan A, et al. Development and initial validation of the Cat HEalth and Wellbeing (CHEW) Questionnaire: a generic health-related quality

of life instrument for cats. *J Feline Med Surg* 2016; 18: 689–701.

12. Bijsmans ES, Jepson RE, Syme HM, et al. Psychometric Validation of a General Health Quality of Life Tool for Cats Used to Compare Healthy Cats and Cats with Chronic Kidney Disease. *J Vet Intern Med* 2016; 30: 183–191.

13 Tatlock S, Gober M, Williamson N, et al. Development and preliminary psychometric evaluation of an owner-completed measure of feline quality of life. *The Veterinary Journal* 2017; 228: 22-32.

14. Osse BHP, Vernooij-Dassen MJFJ, Schadé E, et al. A practical instrument to explore patients' needs in palliative care: the Problems and Needs in Palliative Care questionnaire — short version. *Palliat Med* 2007; 21: 391–399.

15. Reid J, Wiseman-Orr ML, Scott EM, et al. Development, validation and reliability of a web-based questionnaire to measure health-related quality of life in dogs. *J Small Anim Pract @BULLET J Small Anim Pract* 2013; 54: 227–233.

16. Las Hayas C, Quintana JM, Padierna JA, et al. Use of rasch methodology to develop a short version of the Health Related Quality of life for Eating Disorders questionnaire: a prospective study. *Health Qual Life Outcomes* 2010; 8: 29.

17. Wiseman-Orr ML, Scott EM, Reid J, et al. 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. *Am J Vet Res* 2006; 67: 1826–1836.

18. Johnston AL, File SE. Sex differences in animal tests of anxiety. *Physiol Behav* 1991; 49: 245–250.

19. Wiseman-Orr ML, Nolan AM, Reid J, et al. Development of a questionnaire to measure the effects of chronic pain on health-related quality of life in dogs. *Am J Vet Res* 2004; 65: 1077–1084.

20. Reid J, Wiseman-Orr ML and Scott EM. Shortening of an existing online health-related quality of life (HRQL) instrumenfor dogs. *JSAP* Epub ahead of print 2017 doi: 10.1111/jsap.12772

21. Oppenheim AN. *Questionnaire design, interviewing and attitude measurement*. 2nd ed Bloomsbury Publishing; 2000

22. Foddy WH. *Constructing questions for interviews and questionnaires : theory and practice in social research*. Cambridge University Press, 1993.

23.. Glaser B. Grounded theory methodology. Introd Qual Res Psychol. 2013;69–82. *Introd Qual Res Psychol* 2013; 3: 69–82.

24. Lynn MR. Determination and quantification of content validity. *Nurs Res*; 35: 382–5.

25. Schelven AR van, Dikken J, Sillekens LGM, et al. Content Validation of the Dutch Version of the 'Older Patients in Acute Care Survey', an Instrument to Measure the Attitude of Hospital Nurses towards Older Patients. *Int J Clin Med* 2015; 6: 7–18.

26. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol*

Bull 1979; 86: 420–8.

27. Spofford N, Lefebvre SL, McCune S, et al. Should the veterinary profession invest in developing methods to assess quality of life in healthy dogs and cats? *J Am Vet Med Assoc* 2013; 243: 952–956.

28. Melzack R. The McGill Pain Questionnaire: major properties and scoring methods. *Pain* 1975; 1: 277–99.

29. Juniper EF, Guyatt GH, Jaeschke R. How to develop and validate a new health-related quality of life instrument. *Qual Life Pharmacoeconomics Clin Trials* 1996; 49–56.

30. Armstrong FD, Toledano SR, Miloslavich K, et al. The Miami pediatric quality of life questionnaire: parent scale. *Int J Cancer Suppl* 1999; 12: 11–7.

31. Favrot C, Linek M, Mueller R, et al. Development of a questionnaire to assess the impact of atopic dermatitis on health-related quality of life of affected dogs and their owners. *Vet Dermatol* 2010; 21: 64–70.

32. Walton MB, Cowderoy E, Lascelles D, et al. Evaluation of Construct and Criterion Validity for the 'Liverpool Osteoarthritis in Dogs' (LOAD) Clinical Metrology Instrument and Comparison to Two Other Instruments. *PLoS One* 2013; 8: e58125.

33. Polit DF, Beck CT. The content validity index: Are you sure you know what's being reported? critique and recommendations. *Res Nurs Health* 2006; 29: 489–497.

34. Shevlin M, Miles JNV. Effects of sample size, model specification and factor loadings on the GFI in confirmatory factor analysis. *Pers Individ Dif* 1998; 25: 85–90.

35. MacCallum RC, Widaman KF, Zhang S, et al. Sample size in factor analysis. *Psychol Methods* 1999; 4: 84–99.

36. Coste J, Bouée S, Ecosse E, et al. Methodological issues in determining the dimensionality of composite health measures using principal component analysis: case illustration and suggestions for practice. *Qual Life Res* 2005; 14: 641–54.

37. 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: 1293–300.

3**8.** Manificat S. A new instrument to evaluate infant quality of life. Qual Life Newsl. MAPI RESEARCH INSTITUTE; 1999;7–8. *Qual Life Newsl MAPI Res Inst* 1999; 7–8.

39. Stallard P, Williams L, Velleman R, et al. The development and evaluation of the pain indicator for communicatively impaired children (PICIC). *Pain* 2002; 98:

40 Mukaka MM. A guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal* 2012;24(3):69-71.

41. Rosner B. *Fundamentals of biostatistics*. 8th ed. Brooks Cole, 2015.

Figure 1: Summary of study design for developing a generic HRQL instrument for cats Scores are presented as a profile of 3 HRQL domains.

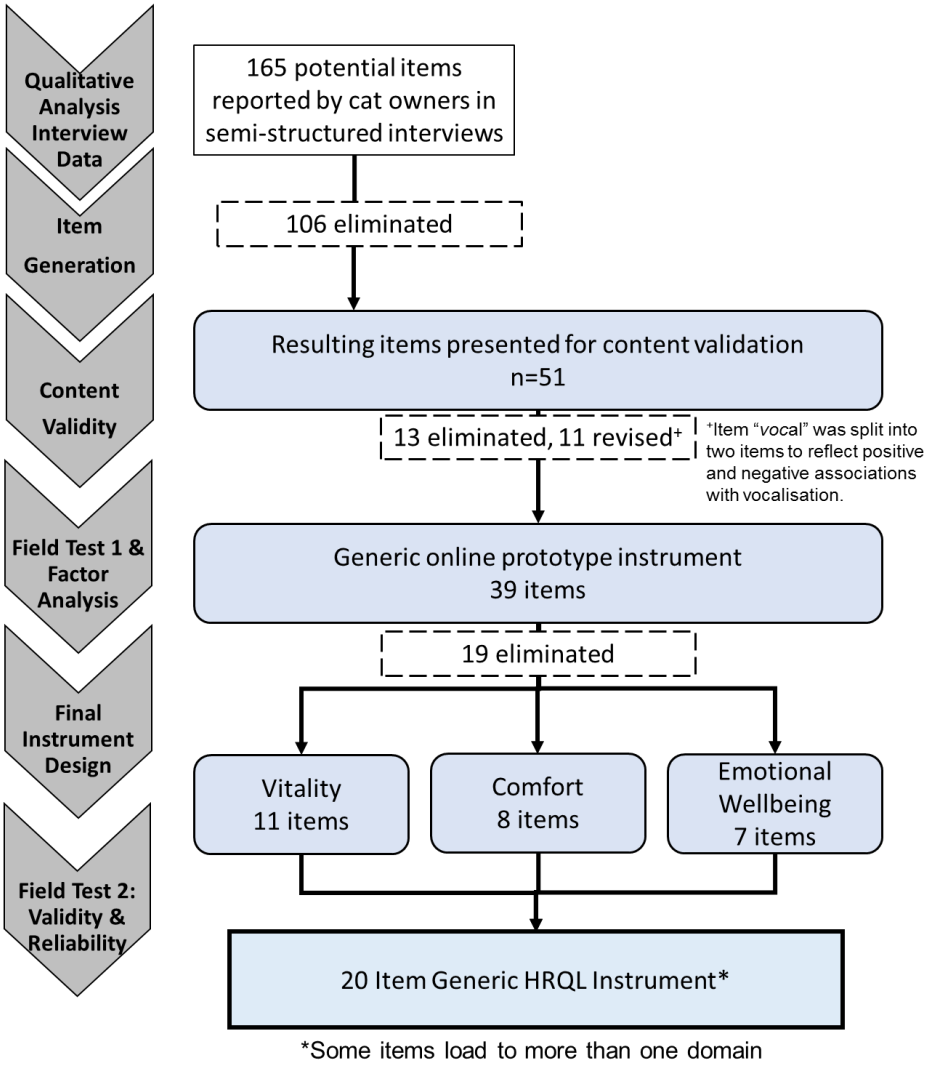

Figure 2: Examples of dotplots for items that were excluded (A) and retained (B) on the basis of their discriminatory potential as assessed by the research team. The x-axis represents the response values selected by owners of healthy (red square) and unhealthy (blue circle) cats.

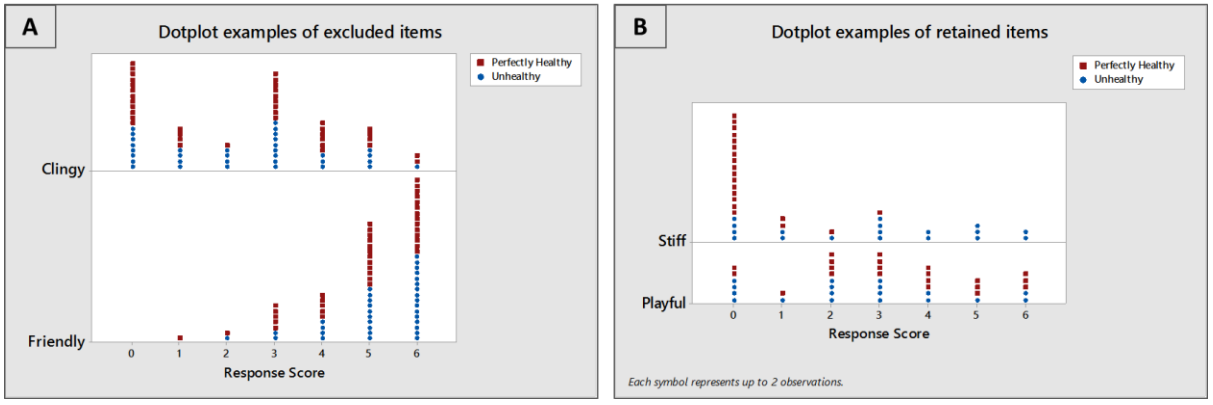

Figure 3: Box plots of scores for three domains of HRQL (Vitality, Comfort and Emotional Wellbeing [EWB]) generated by owners of 36 healthy control cats and 58 sick cats using the 20 item web-based generic HRQL instrument. Each blue box represents the inter-quartile range (middle 50% of scores lie) of scores obtained of the group with the line in the middle representing the median score.

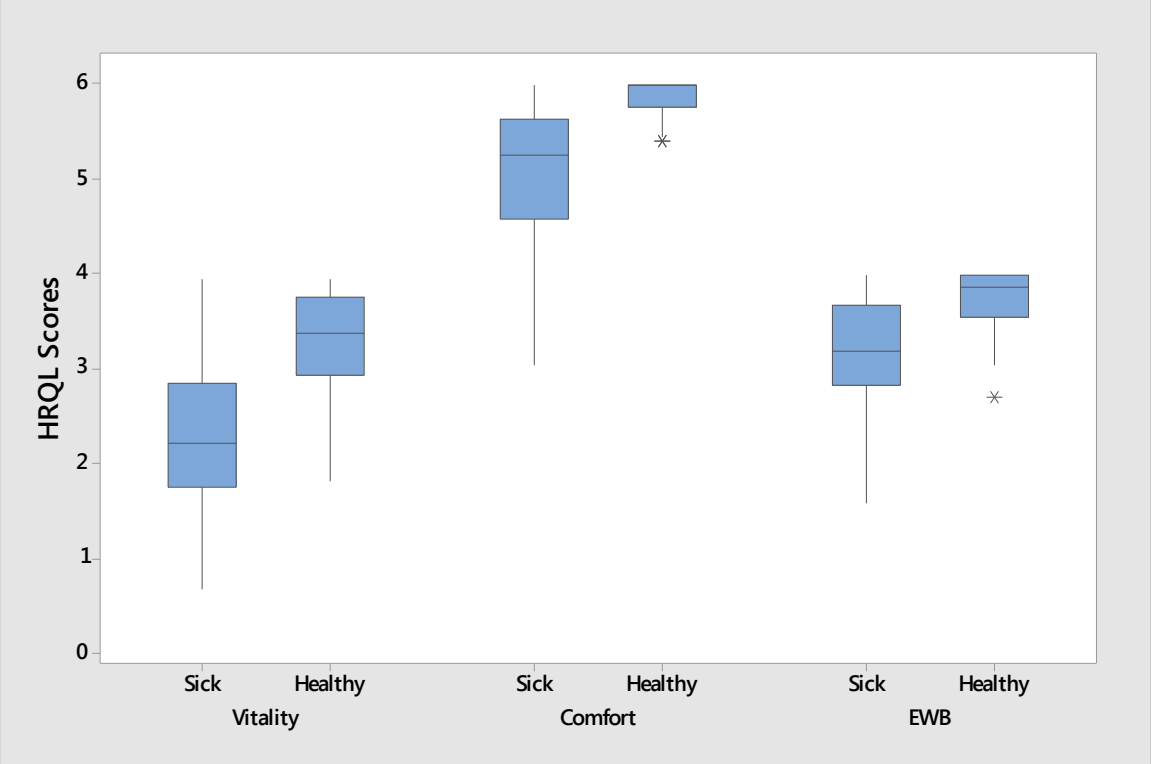

Table 1: Details of owners and cats involved in different stages of instrument development and validation including semi-structured interviews, Field test 1 (item reduction and scoring algorithm generation) and Field test 2 (validity and reliability testing). Demographics of cats include age, health status and presenting conditions, sex, and breed. Misclassification between owner impression and clinician report of health status is reported for each study.

Semi-structured interview	Field Test 1- online testing of prototype instrument	Field Test 2 –testing for validity and reliability
•Cat owners (18, 5 Male, 13 female) •Single & multi-cat households (2 to 4 cats) •Cat age range 1.75 to 21yo •19 healthy & 10 sick (4 Osteoarthritis (OA) and hyperthyroidism, 2 OA and chronic kidney disease, 4 OA only)	•71 single owner assessments •30 healthy, 41 sick cats •43 male, 28 female •Mean age healthy 6.5 (range 0.3 to 16.5) •Mean age sick 11.5y (range 1.2 to 19.5y) •66% domestic short hair, 34% pure breed (i.e. Maine Coon, Persian, Siamese, etc.) •95% comorbidities (Conditions in Table 5) •29% of owners of unhealthy cats misclassified their cats as being healthy- in contrast to clinical assessment	•94 single owner assessments •36 healthy, 58 sick cats •48 male, 46 female •48 repeat assessments •Mean age healthy 4.6y (range 1 to 10y) •Mean age sick 11.7y (range 1.1 to 19.9) •86% domestic short hair, 14% pure breed •72% comorbidities (Conditions in Table 5) •26% of owners of unhealthy cats misclassified their cats as being healthy- in contrast to clinical assessment

Table 2: Examples of questions asked of owners during semi-structured interviews.

Examples of questions asked of cat owners
Please describe your cats' daily routine.
Can you tell how your cat is feeling, and if so, describe how?
How do you know when your cat is unwell? Health? Feeling happy?
How did you first know your cat was unwell? Were there any behavioural changes you noticed specifically?
How do you monitor that the disease is getting worse?
How do you know that treatment is working? Or not working?
What areas of your cats' life are most impacted by the condition?

Table 3: Number of potential items reported by cat owners that were eliminated or revised throughout development following review by the research team (CN, LW, MS, AN, JR) including the rationale and select examples.

Potential items	Rationale	Examples
21 eliminated	Explicitly described underlying personality traits	*"Gentle", "mischievous", "bold"*
22 eliminated	Specific to one disease	*"Yowling", "needing manual evacuation", "night howling"*
	Clinical potential items	*"Doesn't like joint manipulation at the vet", "muscle wastage", weight loss/gain"*
	Specific to individual cat or not easily recognisable	*"Whiskers fanned out", "easy to give medication to when she's feeling well", "head butts"*
	Not relevant to measuring HRQL	*"Runs away after difficulty giving him his medication", "cloudy eyes", "bright eyes"*
63 eliminated	True synonyms/more appropriate descriptor commonly used	*"Friendly"* was most commonly reported and synonym of *"sociable", "follows me around" and "come to greet you".*
	More appropriate descriptor available	*"Content"* was chosen instead of *"chilled"* as it was believed more appropriate for a wider audience
	Another descriptor would adequately capture a behaviour over one that is "too specific"	"*Interested in his/her food*" was retained, covering "*loss of appetite*", "*enjoying food*", "*less hungry*"
11 revised	Revised to improve clarity/ readability	*"Doesn't go out in winter anymore" was revised to "going out in cold weather"; "getting up and down the stairs" to "managing getting up and down the stairs"*

Table 4: Twenty items that make up the feline HRQL scale and their response options

1 Please tell us how well this word describes (cat name) as he is today: Active

Couldn't be more active - Not at all active: 6-0

2 Please tell us how well this word describes (cat name) as he is today: Unsteady

Couldn't be more unsteady - Not at all unsteady: 6-0

3 Please tell us how well this word describes (cat name) as he is today: Energetic

Couldn't be more energetic - Not at all energetic: 6-0

4 Please tell us how well this word describes (cat name) as he is today: Comfortable

Couldn't be more comfortable - Not at all comfortable: 6-0

5 Please tell us how well this word describes (cat name) as he is today: Lethargic

Couldn't be more lethargic - Not at all lethargic: 6-0

6 Please tell us how well this word describes (cat name) as he is today: Showing hunting behaviour

Couldn't be showing hunting behaviour more - Not at all showing hunting behaviour: 6-0

7 Please tell us how well this word describes (cat name) as he is today: Lively

Couldn't be more lively - Not at all lively: 6-0

8 Please tell us how well this word describes (cat name) as he is today: Alert

Couldn't be more alert - Not at all alert: 6-0

9 Please tell us how well this word describes (cat name) as he is today: Sore

Couldn't be more sore - Not at all sore: 6-0

10 Please tell us how well this word describes (cat name) as he is today: Content

Couldn't be more content - Not at all content: 6-0

11 Please tell us how well this word describes (cat name) as he is today: Playful

Couldn't be more playful - Not at all playful: 6-0

12 Please tell us how well this word describes (cat name) as he is today: Uncomfortable

Couldn't be more uncomfortable- Not at all uncomfortable: 6-0

13 Please tell us how well this word describes (cat name) as he is today: Enjoying the things he usually does

Couldn't be enjoying the things he usually does more - Not at all enjoying the things he usually does: 6-0

14 Please tell us how well this word describes (cat name) as he is today: Jumping or climbing up/down as usual

Jumping or climbing up/down as usual- Not jumping or climbing up/down as usual: 6-0

15 Please tell us how well this word describes (cat name) as he is today: Exploring

Couldn't be exploring more - Not at all exploring: 6-0

16 Please tell us how well this word describes (cat name) as he is today: Feeling himself

Couldn't be feeling himself more - Not at all feeling himself: 6-0

17 Please tell us how well this word describes (cat name) as he is today: Stiff

Couldn't be more stiff - Not at all stiff: 6-0

18 Please tell us how well this word describes (cat name) as he is today: Happy

Couldn't be more happy- Not at all happy: 6-0

19 Please tell us how well this word describes (cat name) as he is today: Inquisitive

Couldn't be more inquisitive- Not at all inquisitive: 6-0

20 Please tell us how well this word describes (cat name) as he is today: Slow

Couldn't be more slow- Not at all slow: 6-0

Table 5: Conditions reported by clinicians for each assessment completed for 41 cats from Field test1 and the 58 cats from Field test 2 that in the clinician's opinion were NOT perfectly healthy. Comorbid conditions were reported in 95% and 72% of cases in Field tests 1 and 2 respectively.

Presenting Conditions	Field Test 1	Field Test 2
Degenerative joint disease	23	33
Obesity	8	11
Painful cancer	0	1
Non - painful cancer	3	4
Chronic skin disease	7	1
Chronic medical condition	28	21
Cardiac disease	11	4
Neurological disease	2	1
Chronic ear disease	3	2
Chronic dental disease	21	9
Chronic kidney disease	20	16
Hyperthyroidism	6	9
Chronic lower urinary tract disease	2	6
Cat flu	2	3
Chronic gastrointestinal disease	10	5
Previous physical trauma	4	3
Under-weight	12	21

Other	7*	2*

*Other conditions included: Field Test 1, cognitive decline (3), cancer in remission (2), diabetes (2), proliferative gum disease- a symptomatic, liver disease, and otitis externa right ear & scratches to pinna; Field Test 2, hypertension (2).

589

590

591

592

593 **Table 6:** Descriptive statistics and Mann-Whitney test results comparing the scores of healthy
594 and sick cats for each of the three domains (Vitality, Comfort and Emotional Wellbeing
595 [EWB]).

596
597
598
599

Domain	Number of cats	Mean +/- SD	Interquartile range	Median	Mann-Whitney Difference in median (healthy-unwell)	P-value	95% confidence interval
Vitality							
Healthy	36	3.30+/-0.53	0.84	3.37	1.04	<0.001	(0.71, 1.32)
Sick	58	2.32+/-0.75	1.08	2.23			
Comfort							
Healthy	36	5.90+/-0.20	0.23	6.00	0.63	<0.001	(0.40, 0.98)
Sick	58	5.06+/-0.73	1.06	5.25			
Emotional Wellbeing							
Healthy	36	3.73+/-0.32	0.45	3.87	0.56	<0.001	(0.33, 0.77)
Sick	58	3.15+/-0.63	0.83				

600

601

602

603

604
605
606
607