



Noble, C.E., Wiseman-Orr, M.L., Scott, E.M. , Nolan, A.M. and Reid, J. (2018) Development, initial validation and reliability testing of a web-based, generic feline health-related quality of life instrument. *Journal of Feline Medicine and Surgery*, (doi:[10.1177/1098612X18758176](https://doi.org/10.1177/1098612X18758176))

This is the author's final accepted version.

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

<http://eprints.gla.ac.uk/157133/>

Deposited on: 12 February 2018

Enlighten – Research publications by members of the University of Glasgow
<http://eprints.gla.ac.uk>

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

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

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

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

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

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

12

13 **Corresponding author:**

14 *Prof Reid, BVMS PhD DVA DipECVAA MRCA MRCVS*

15 *jacky.reid@newmetrica.com*

16 University of Glasgow
17 CIDS Incubator, Rm 203
18 Thomson Building
19 University Avenue
20 Glasgow, G12 8QQ

21 **+44 (0) 7876683262**

22

23 **Abstract**

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

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

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

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

46

47 **Introduction**

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

55 Instruments to measure HRQL in companion animals consist of questions for the owner,
56 who is well placed to report upon the subtle changes in behaviour, attitude and demeanour
57 that occur with chronic disease. While several feline disease-specific instruments exist (5–9),
58 to date, no validated generic HRQL instrument exists for the purpose of comparing
59 treatments or disease states. (10) A generic instrument (CHEW), based on owner perceived
60 health status has been reported (11), but not validated in sick cats. Similarly CatQol
61 (Bijsmans) focuses on general health, eating, behavior and management, but has been
62 validated in cats with chronic kidney disease (CKD) only. (12) More recently Tatlock et al
63 (2017) have described an owner reported feline quality of life scale for healthy cats. (13)

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

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

91 A reliable instrument will produce the same score when an unchanging subject is measured
92 at two time points by the same observer (repeatability/intra-rater reliability), or when two
93 people measure the same subject at one time (reproducibility/inter-rater reliability). (2)

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

101

102 **Materials and methods**

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

105

106 ***Item generation and initial selection***

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

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

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

123 ***Content validation***

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

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

136 ***Prototype construction and pre-testing***

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

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

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

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

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

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

156 To establish evidence for factorial validity, and to determine a scoring algorithm for the
157 instrument, a factor analysis (principal components method with a varimax rotation) was
158 performed with remaining items. The scores attributed to each item by the owners were
159 used for the factor analysis. The analysis allocates each item to a factor with a loading (0 – 1)
160 which determines the closeness of its relationship to the factor. Resulting loadings were
161 sorted, and any item with a loading <0.4 was excluded. A scree test and the Kaiser criterion
162 were used to identify the optimum number of factors and the interpretability of a range of
163 factor models was examined. Factors were interpreted based on how those items loading
164 onto a particular factor were related, and a factor model was chosen that accounted for an
165 acceptable amount of the variability in the data, was readily interpretable, and did not
166 include any factors containing only one or two items. An algorithm, based on the item–
167 factor associations of the selected factor model, was derived in order to generate a domain
168 score for each of the resultant factors/domains.

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

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

174 Owners of healthy and sick cats, grouped according to the clinical judgement of the
175 consulting clinician, completed at least one assessment. A number of owners of healthy cats
176 completed 2 assessments, 2 weeks apart, and test–retest reliability was assessed using the
177 intraclass correlation coefficient (ICC). A one-way random model was assumed where the
178 subjects (cats) are assumed random. (26)

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

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

190

191 **Results**

192 An overview of the development process is shown in Figure 1

193 ***Item generation and initial selection***

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

199 ***Content validation***

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

204 ***Prototype construction and pre-testing***

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

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

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

214 Using the online prototype instrument for Field Test 1, 71 owner and clinician assessments
215 from UGSAH, 5 general practices and 1 feline specialist practice were completed over a
216 period of 5 months for 30 healthy cats and 41 cats diagnosed with a chronic condition
217 expected to impact QOL (Table 1). Ninety-five percent of cats presented with 1 to 6
218 comorbidities (Table 5). Review of dot plots of these item responses suggested that 19 items
219 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

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

250 ***Reliability***

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

255

256 **Discussion**

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

267 19 healthy cats and 10 sick cats belonging to owners recruited as key informants in this
268 study was verified by a veterinary surgeon. Although this number of cats could be
269 considered to be low, interviews were conducted until no new information emerged. In
270 addition, 48 different owners involved in the content validation process were invited to
271 suggest additional items if they felt the collection of items was inadequate. Initial reduction
272 of the 165 items was based on criteria devised by the investigators, an approach considered
273 to be appropriate in human medicine. (28–30)

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

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

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

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

292

293 The remaining 20 items all loaded >0.4 in the FA. Factor loadings of 0.3, 0.5 and 0.7 are
294 generally considered to be low, medium and high respectively (34) with loadings of 0.3
295 deemed to be the minimum consideration level for FA. Increasing the loading threshold to
296 0.4 may have accounted for the fact that no further items were removed at this stage.
297 Furthermore, the fact that the majority of loadings were >0.6 indicates stability of the factor
298 model. (35) Although FA provides any number of factor models for a given data set, there
299 are established methods of identifying how many factors could sensibly be extracted,
300 including the scree test and the Kaiser criterion (36), both considered in this study. The 3
301 factor model adopted here, accounting for 72% of the variability in the owner response
302 data, compares favorably with the canine HRQL (64%)(15) and 72% for the shortened
303 instrument (20), an 11-factor questionnaire designed to measure the behaviour and
304 temperament of pet dogs (57%) (37) and a 4-factor QoL questionnaire regarding infants
305 (38), and confirms factorial validity.

306

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

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

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

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

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

341 In conclusion, the measurement of feline HRQL is becoming more necessary as chronic
342 diseases such as CKD, hyperthyroidism, cognitive decline and OA affect the QOL of an
343 increasing number of ageing cats, and evidence-based medicine requires that robust
344 measures of clinical impact be developed. This study has provided initial evidence for the
345 reliability and validity of a novel generic instrument that measures the affective component
346 of the chronic disease experience. However it is important to emphasise that validity is not
347 determined by a single statistic, but by a body of research that supports the claim that the
348 instrument is valid for particular purposes, with defined populations and in specified
349 contexts. (2) Accordingly, future research will seek to provide such evidence, as well as
350 evidence for the instrument’s responsiveness to clinical change including that following
351 treatment. The instrument is available for clinical use and for clinical trials from NewMetrica
352 (www.newmetrica.com). For further information please contact the corresponding author,
353 jacky.reid@newmetrica.com.

354 **Acknowledgments**

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

358 **Funding**

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

361 **Conflict of Interests**

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

364 **References**

- 365 1. Fayers PM, Machin D. *Quality of life : the assessment, analysis and interpretation of*
366 *patient-reported outcomes*. John Wiley & Sons, 2013.
- 367 2. Streiner DL, Norman GR, Cairney J. *Health Measurement Scales: a practical guide to*
368 *their development and use*. Oxford University Press, USA, 2015.
- 369 3. Abell N, Springer DW, Kamata A. *Developing and Validating Rapid Assessment*
370 *Instruments*. Oxford University Press, 2009.
- 371 4. Brod M, Tesler LE, Christensen TL. Qualitative research and content validity:
372 developing best practices based on science and experience. *Qual Life Res* 2009; 18:
373 1263–1278.
- 374 5. Freeman LM, Rush JE, Oyama MA, et al. Development and evaluation of a
375 questionnaire for assessment of health-related quality of life in cats with cardiac
376 disease. *J Am Vet Med Assoc* 2012; 240: 1188–1193.
- 377 6. Zamprogno H, Hansen BD, Bondell HD, et al. Item generation and design testing of a
378 questionnaire to assess degenerative joint disease–associated pain in cats. *Am J Vet*
379 *Res* 2010; 71: 1417–1424.
- 380 7. Niessen SJM, Powney S, Guitian J, et al. Evaluation of a Quality-of-Life Tool for Cats
381 with Diabetes Mellitus. *J Vet Intern Med* 2010; 24: 1098–1105.
- 382 8. Noli C, Minafò G, Galzerano M. Quality of life of dogs with skin diseases and their
383 owners. Part 1: development and validation of a questionnaire. *Vet Dermatol* 2011;
384 22: 335–343.
- 385 9. Noli C, Borio S, Varina A, et al. Development and validation of a questionnaire to
386 evaluate the Quality of Life of cats with skin disease and their owners, and its use in
387 185 cats with skin disease. *Vet Dermatol* 2016; 27: 247-e58.
- 388 10. Giuffrida MA, Kerrigan SM. Quality of life measurement in prospective studies of
389 cancer treatments in dogs and cats. *J Vet Intern Med* 2014; 28: 1824–9.
- 390 11. Freeman LM, Rodenberg C, Narayanan A, et al. Development and initial validation of
391 the Cat HEalth and Wellbeing (CHEW) Questionnaire: a generic health-related quality

- 392 of life instrument for cats. *J Feline Med Surg* 2016; 18: 689–701.
- 393 12. Bijsmans ES, Jepson RE, Syme HM, et al. Psychometric Validation of a General Health
394 Quality of Life Tool for Cats Used to Compare Healthy Cats and Cats with Chronic
395 Kidney Disease. *J Vet Intern Med* 2016; 30: 183–191.
- 396 13. Tatlock S, Gober M, Williamson N, et al. Development and preliminary psychometric
397 evaluation of an owner-completed measure of feline quality of life. *The Veterinary*
398 *Journal* 2017; 228: 22-32.
- 399 14. Osse BHP, Vernooij-Dassen MJFJ, Schadé E, et al. A practical instrument to explore
400 patients' needs in palliative care: the Problems and Needs in Palliative Care
401 questionnaire — short version. *Palliat Med* 2007; 21: 391–399.
- 402 15. Reid J, Wiseman-Orr ML, Scott EM, et al. Development, validation and reliability of a
403 web-based questionnaire to measure health-related quality of life in dogs. *J Small*
404 *Anim Pract @BULLET J Small Anim Pract* 2013; 54: 227–233.
- 405 16. Las Hayas C, Quintana JM, Padierna JA, et al. Use of rasch methodology to develop a
406 short version of the Health Related Quality of life for Eating Disorders questionnaire:
407 a prospective study. *Health Qual Life Outcomes* 2010; 8: 29.
- 408 17. Wiseman-Orr ML, Scott EM, Reid J, et al. Validation of a structured questionnaire as
409 an instrument to measure chronic pain in dogs on the basis of effects on health-
410 related quality of life. *Am J Vet Res* 2006; 67: 1826–1836.
- 411 18. Johnston AL, File SE. Sex differences in animal tests of anxiety. *Physiol Behav* 1991;
412 49: 245–250.
- 413 19. Wiseman-Orr ML, Nolan AM, Reid J, et al. Development of a questionnaire to
414 measure the effects of chronic pain on health-related quality of life in dogs. *Am J Vet*
415 *Res* 2004; 65: 1077–1084.
- 416 20. Reid J, Wiseman-Orr ML and Scott EM. Shortening of an existing online
417 health-related quality of life (HRQL) instrument for dogs. *JSAP Epub*
418 ahead of print 2017 doi: 10.1111/jsap.12772
419
- 420 21. Oppenheim AN. *Questionnaire design, interviewing and attitude measurement*. 2nd ed
421 Bloomsbury Publishing; 2000
- 422 22. Foddy WH. *Constructing questions for interviews and questionnaires : theory and*
423 *practice in social research*. Cambridge University Press, 1993.
- 424 23.. Glaser B. Grounded theory methodology. *Introd Qual Res Psychol*. 2013;69–82. *Introd*
425 *Qual Res Psychol* 2013; 3: 69–82.
- 426 24. Lynn MR. Determination and quantification of content validity. *Nurs Res*; 35: 382–5.
- 427 25. Schelven AR van, Dikken J, Sillekens LGM, et al. Content Validation of the Dutch
428 Version of the 'Older Patients in Acute Care Survey', an Instrument to Measure the
429 Attitude of Hospital Nurses towards Older Patients. *Int J Clin Med* 2015; 6: 7–18.
- 430 26. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol*

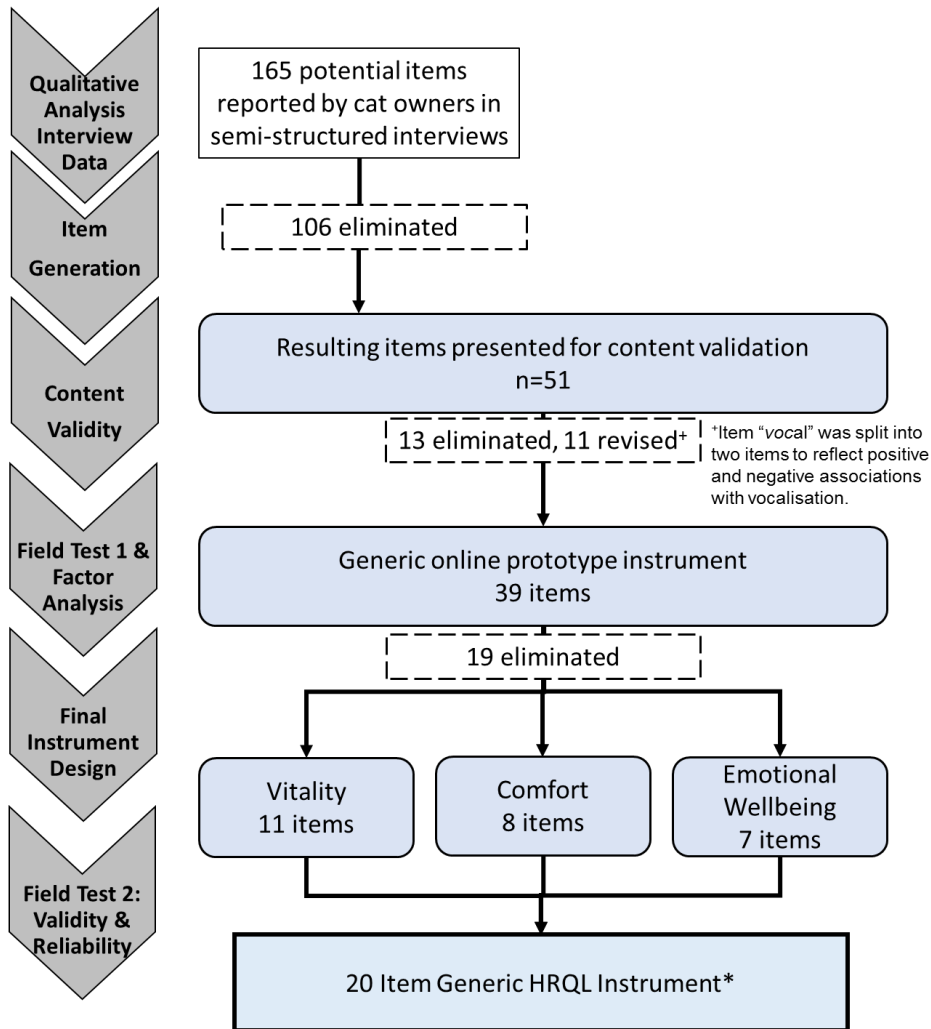
- 431 *Bull* 1979; 86: 420–8.
- 432 27. Spofford N, Lefebvre SL, McCune S, et al. Should the veterinary profession invest in
433 developing methods to assess quality of life in healthy dogs and cats? *J Am Vet Med*
434 *Assoc* 2013; 243: 952–956.
- 435 28. Melzack R. The McGill Pain Questionnaire: major properties and scoring methods.
436 *Pain* 1975; 1: 277–99.
- 437 29. Juniper EF, Guyatt GH, Jaeschke R. How to develop and validate a new health-related
438 quality of life instrument. *Qual Life Pharmacoeconomics Clin Trials* 1996; 49–56.
- 439 30. Armstrong FD, Toledano SR, Miloslavich K, et al. The Miami pediatric quality of life
440 questionnaire: parent scale. *Int J Cancer Suppl* 1999; 12: 11–7.
- 441 31. Favrot C, Linek M, Mueller R, et al. Development of a questionnaire to assess the
442 impact of atopic dermatitis on health-related quality of life of affected dogs and their
443 owners. *Vet Dermatol* 2010; 21: 64–70.
- 444 32. Walton MB, Cowderoy E, Lascelles D, et al. Evaluation of Construct and Criterion
445 Validity for the ‘Liverpool Osteoarthritis in Dogs’ (LOAD) Clinical Metrology
446 Instrument and Comparison to Two Other Instruments. *PLoS One* 2013; 8: e58125.
- 447 33. Polit DF, Beck CT. The content validity index: Are you sure you know what’s being
448 reported? critique and recommendations. *Res Nurs Health* 2006; 29: 489–497.
- 449 34. Shevlin M, Miles JNV. Effects of sample size, model specification and factor loadings
450 on the GFI in confirmatory factor analysis. *Pers Individ Dif* 1998; 25: 85–90.
- 451 35. MacCallum RC, Widaman KF, Zhang S, et al. Sample size in factor analysis. *Psychol*
452 *Methods* 1999; 4: 84–99.
- 453 36. Coste J, Bouée S, Ecosse E, et al. Methodological issues in determining the
454 dimensionality of composite health measures using principal component analysis:
455 case illustration and suggestions for practice. *Qual Life Res* 2005; 14: 641–54.
- 456 37. Hsu Y, Serpell JA. Development and validation of a questionnaire for measuring
457 behavior and temperament traits in pet dogs. *J Am Vet Med Assoc* 2003; 223: 1293–
458 300.
- 459 38. Manificat S. A new instrument to evaluate infant quality of life. *Qual Life Newsl. MAPI*
460 *RESEARCH INSTITUTE*; 1999;7–8. *Qual Life Newsl MAPI Res Inst* 1999; 7–8.
- 461 39. Stallard P, Williams L, Velleman R, et al. The development and evaluation of the pain
462 indicator for communicatively impaired children (PICIC). *Pain* 2002; 98:
- 463 40 Mukaka MM. A guide to appropriate use of correlation coefficient in medical
464 research. *Malawi Medical Journal* 2012;24(3):69-71.
- 465 41. Rosner B. *Fundamentals of biostatistics*. 8th ed. Brooks Cole, 2015.

466

467

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

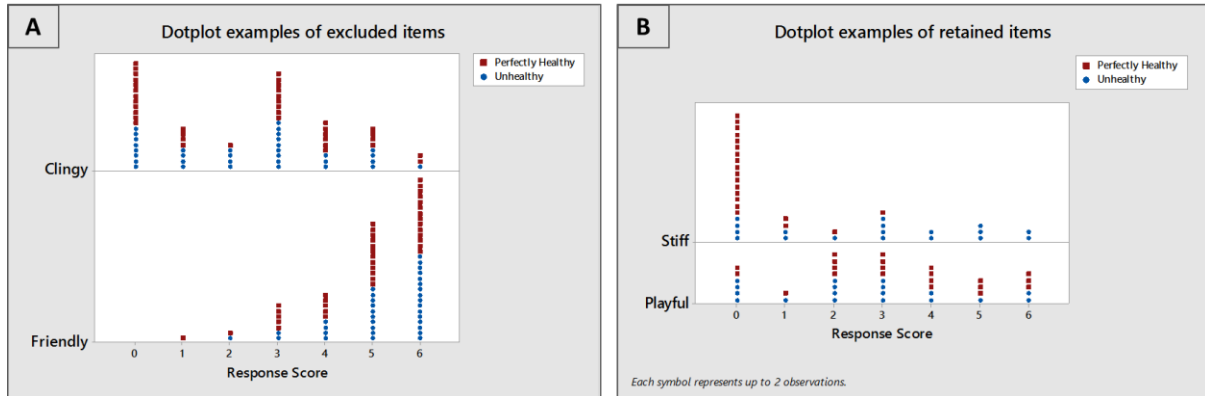
470
 471



*Some items load to more than one domain

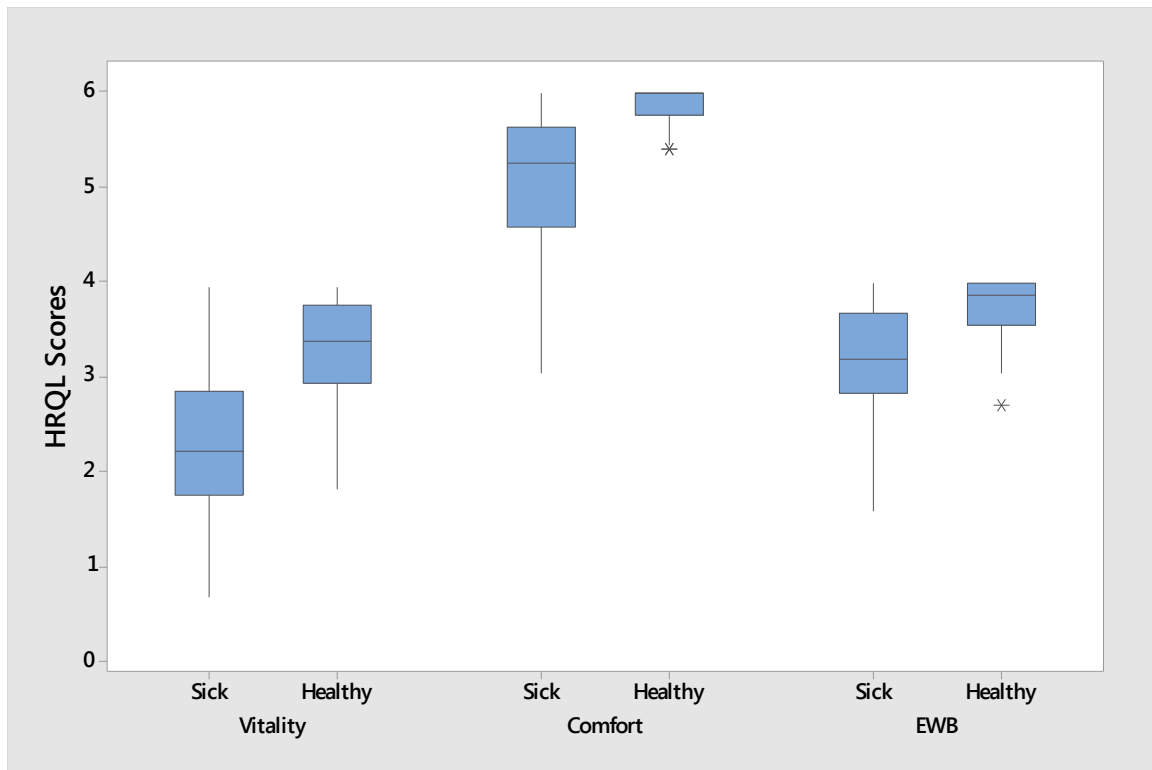
472
 473
 474
 475
 476
 477
 478
 479
 480
 481

482 **Figure 2:** Examples of dotplots for items that were excluded (A) and retained (B) on the basis
 483 of their discriminatory potential as assessed by the research team. The x-axis represents the
 484 response values selected by owners of healthy (red square) and unhealthy (blue circle) cats.
 485
 486



487
 488
 489
 490
 491
 492
 493
 494
 495

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.



496
 497
 498
 499
 500
 501

502 **Table 1:** Details of owners and cats involved in different stages of instrument development
 503 and validation including semi-structured interviews, Field test 1 (item reduction and scoring
 504 algorithm generation) and Field test 2 (validity and reliability testing). Demographics of cats
 505 include age, health status and presenting conditions, sex, and breed. Misclassification
 506 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
<ul style="list-style-type: none"> •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) 	<ul style="list-style-type: none"> •71 single owner assessments <ul style="list-style-type: none"> •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 	<ul style="list-style-type: none"> •94 single owner assessments <ul style="list-style-type: none"> •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

507

508

509

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

511

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?

512

513

514

515

516

517

518

519

520

521

522

523

524

525

526

527

528
 529
 530
 531
 532
 533

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"

534
 535
 536
 537
 538
 539
 540

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

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

542 **2 Please tell us how well this word describes (cat name) as he is today: Unsteady**

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

544 **3 Please tell us how well this word describes (cat name) as he is today: Energetic**

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

546 **4 Please tell us how well this word describes (cat name) as he is today: Comfortable**

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

548 **5 Please tell us how well this word describes (cat name) as he is today: Lethargic**

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

550 **6 Please tell us how well this word describes (cat name) as he is today: Showing hunting**

551 **behaviour**

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

553 **7 Please tell us how well this word describes (cat name) as he is today: Lively**

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

555 **8 Please tell us how well this word describes (cat name) as he is today: Alert**

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

557 **9 Please tell us how well this word describes (cat name) as he is today: Sore**

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

559 **10 Please tell us how well this word describes (cat name) as he is today: Content**

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

561 **11 Please tell us how well this word describes (cat name) as he is today: Playful**

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

563 **12 Please tell us how well this word describes (cat name) as he is today: Uncomfortable**

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

565 **13 Please tell us how well this word describes (cat name) as he is today: Enjoying the**

566 **things he usually does**

567 Couldn't be enjoying the things he usually does more - Not at all enjoying the things he

568 usually does: 6-0

569 **14 Please tell us how well this word describes (cat name) as he is today: Jumping or**

570 **climbing up/down as usual**

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

572 **15 Please tell us how well this word describes (cat name) as he is today: Exploring**
573 Couldn't be exploring more - Not at all exploring: 6-0
574 **16 Please tell us how well this word describes (cat name) as he is today: Feeling himself**
575 Couldn't be feeling himself more - Not at all feeling himself: 6-0
576 **17 Please tell us how well this word describes (cat name) as he is today: Stiff**
577 Couldn't be more stiff - Not at all stiff: 6-0
578 **18 Please tell us how well this word describes (cat name) as he is today: Happy**
579 Couldn't be more happy- Not at all happy: 6-0
580 **19 Please tell us how well this word describes (cat name) as he is today: Inquisitive**
581 Couldn't be more inquisitive- Not at all inquisitive: 6-0
582 **20 Please tell us how well this word describes (cat name) as he is today: Slow**
583 Couldn't be more slow- Not at all slow: 6-0

584

585 **Table 5:** Conditions reported by clinicians for each assessment completed for 41 cats from
586 Field test1 and the 58 cats from Field test 2 that in the clinician’s opinion were NOT
587 perfectly healthy. Comorbid conditions were reported in 95% and 72% of cases in Field tests
588 1 and 2 respectively.

Presenting Conditions	Field Test	Field Test
	1	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

603

604

605

606

607