**Introduction**

Clinical metrology instruments (CMIs), also called owner reported outcome measures (OROMs), are a sequence of questions which are scored based on owners’ observations regarding their pets’ health status. The cumulative score provides the clinician with information on limb function, pain, and their quality of life. Repeated use can be used to monitor disease progression and assess response to medical/surgical interventions.1-3 Whilst kinetic and kinematic gait analysis are considered by many as the gold standard of quantitating canine limb function,4-6 use of validated CMIs is cost-effective, reliable, less time-consuming and does not require specialised equipment. In addition, CMIs aim to capture the overall clinical picture as perceived by the observer and can be used for cases that present with single or multi-limb lameness. On the contrary, objective gait analysis is best used for assessment of single-limb lameness as symmetry index values may be unreliable when both of the thoracic or pelvic limbs of a dog are abnormal.7

Several CMIs have been reported for evaluation of canine osteoarthritis,1,3,8-12 with the Liverpool Osteoarthritis in Dogs (LOAD) and the Canine Brief Pain Inventory (CBPI) questionnaires being amongst the most frequently used. Although the CBPI is a validated CMI, a study has shown that its two-factor structure is not ideally suited to measure owner-perceived pain related to osteoarthritis.13 Both the aforementioned CMIs have only been validated for contemporaneous use.11, 14 For retrospective studies, contemporaneous CMI data may not be available so retrospective collection of CMI data may provide a solution to complete missing data sets. In such a scenario, owners would be asked to complete a CMI recalling their dog’s pre-intervention status. CMIs have been used to retrospectively complete preoperative data sets for total hip replacement (THR), total elbow replacement, cruciate disease and surgical correction of angular limb deformity in dogs.15-19 A study using this methodology may be susceptible to recall bias.20 Recall bias is a particular concern in studies where data is collected retrospectively and is defined as a systematic error in the accuracy or completeness of the recollections by a study participant regarding events or experiences from the past.21, 22 This bias could lead to either overestimation or underestimation of the pre-intervention status.23-25, 27-30 In comparable human research, questionnaires that were completed retrospectively by patients following lumbar spinal or total joint replacement surgery, showed that patients overestimate their preoperative status. 23, 24 Overestimation of preoperative status could inaccurately improve intervention success rates and reduce the accuracy and even the validity of the conclusions made.23, 25 The susceptibility of the LOAD and CBPI CMIs to recall bias is unknown. Therefore, the primary objective of this study was to determine if owners can accurately recall their dog’s pre-treatment status using the LOAD and CBPI questionnaires at predefined time points following their first consultation. A secondary objective of the study was to identify factors impacting owners’ recollection of their dogs’ pre-treatment status. We hypothesised that the agreement between the actual and recalled pre-treatment scores would be poor, based on previous human data.

**Materials and Methods**

**Inclusion Criteria**

Client-owned dogs that presented to the XX XXXXX XXXX XXXXX, XXXXX X XXXXX orthopaedic clinic for investigation of lameness between April 2018 and February 2020 were enrolled in the study. Dogs presenting with an acute traumatic injury and/or dogs with incomplete records were excluded. The study was approved by the XXXXX XXXX XXX Research Ethics Committee (VREC577) and owners were provided an information sheet outlining the study and a written consent form prior to enrolment.

**LOAD and CBPI questionnaires**

The LOAD questionnaire is a 13-item instrument. Each item is scored from 0-4 and the sum of all 13 items is used to generate a final instrument score out of a maximum of 52.8 Patients were also stratified as mildly (0-10/52), moderately (11-20/52), severely (21-30/52) or extremely (31-52/52) affected based on numerical scoring. 26

The CBPI questionnaire is a two-part instrument.14 The first section calculates the Pain Severity Score (PSS) and includes four items scored on an 11-point scale (0-10). The second section calculates the Pain Interference Score (PIS) and includes six items scored on an 11-point scale. The sum on each section was used to generate a PSS score out of 40 and a PIS score out of 60. Overall quality of life (QOL) was rated on a 5-point categorical scale from poor to excellent.

**Questionnaire completion timelines**

Each owner was asked to complete the LOAD and CBPI questionnaires at the time of their dog’s initial presentation to the hospital (T0). The LOAD and CBPI questionnaires were completed again at two (T1), six (T2), and 12 months (T3) after the initial presentation. At these timepoints, owners were asked to complete the questionnaire by trying to recall their dog’s status at T0. When dogs were reassessed at the hospital at T1, paper copies of the CMIs were completed during that visit. When dogs were not scheduled to return for post-treatment evaluation, questionnaires were completed via telephone interview by XX and XX.

**Data collection**

Data collected at T0 included age, gender, breed, weight, affected limb (thoracic or pelvic limb), subjective gait analysis (assessed on a scale from 0 to 1031 and graded as mild (0-3), moderate (4-7) severe (8-10)31), diagnosis and type of management advised (surgical or conservative). LOAD scores, CBPI PSS and PIS scores and CBPI QOL were collected at T0, T1, T2, T3. A LOAD categorical score (mild, moderate, severe, extreme) was also given at the three different time points, deriving from the numerical scores. The number of days at collection of CMI scores at T1, T2 and T3 time points were recorded. Cases where only one of the two questionnaires were completed at T0 were not excluded; for these cases owners completed the same questionnaire at T1, T2 and T3.

**Statistical analysis**

The enrolment of 186 dogs was calculated to provide a study power of 80% assuming moderate agreement (0.6) as assessed by intraclass correlation coefficient (ICC) between scores of 4 observations and with a lower 95% confidence interval for the ICC of no less than 0.5. Data normality was assessed using the Kolmogorov-Smirnov and Shapiro-Wilk tests; normally distributed data are presented as mean ± SD and non-normally distributed data as median and range. The Cohen’s weighted kappa statistic (κw) was used to measure the agreement of LOADcategorical scoring between T0 and T1, T0 and T2, T0 and T3. A κw coefficient of <0.2 indicates poor agreement, 0.2-0.4 fair agreement, 0.41-0.6 moderate agreement, 0.61-0.8 good agreement and >0.8 excellent agreement.32 The agreement of LOAD, CBPI PSS and PIS scores between T0 and T1, T0 and T2, T0 and T3 was assessed using the two-way mixed effects ICC and its 95% confidence intervals (95% CI). ICC of <0.5 indicates poor agreement, 0.5-0.75 moderate agreement, 0.75-0.9 good agreement and >0.9 excellent agreement.33 The Wilcoxon signed rank test was used to assess the difference between paired LOADand CBPI scores between T0 and T1, T0 and T2 and T0 and T3. Multilevel logistic regression was performed to identify factors associated with the absolute difference between initial and subsequent LOAD scores. Within dog clustering of LOAD scores was accounted for as a random intercept term in these two-level models. Any variable with a potential association with the difference in LOAD scores (*p*-value <0.3) was considered for inclusion into the final multivariable model; for any correlated variables (correlation coefficient >0.7), only the variable with the lowest *p*-value of the pair was included. The multivariable model was constructed with a manual backwards stepwise approach with retention of variables with Wald *p*-values <0.05. A *p* value <0.05 was considered statistically significant. Statistical analysis was carried out using the statistical software programs SPSS 25.0 (SPSS Inc, Chicago, Illinois, USA) and MLwiN Version 3.02 (Centre for Multilevel Modelling, university of Bristol, UK).

**Results**

**Animal and Clinical Data**

Of the 257 dogs initially enrolled on the study, 40 were excluded due to incomplete questionnaires at T0, euthanasia, re-homing, or owner withdrawal at T1. Overall, 217 dogs were included for statistical analysis. Of the 77 breeds represented, the most common were mixed-breed dogs (54) followed by the Labrador Retriever (37), English Springer Spaniel (14), Border Collie (13), German Shepherd Dogs (9), West Highland White Terriers (7) and five each of Cocker Spaniels, Golden Retrievers and Staffordshire Bull Terriers. There were 101 female dogs (84 neutered) and 126 male dogs (82 neutered). The mean age was 4.8 ± 3.2 years and the dogs’ mean weight was 23.7 ± 12.2 kg. The thoracic or pelvic limb was affected in 62 and 136 dogs respectively and 20 dogs presented with multi-limb lameness. Lameness was graded as mild in 92 dogs, moderate in 61 dogs, severe in 27 dogs; six dogs exhibited skipping lameness. Conditions were stratified into eight groups: cranial cruciate ligament disease (n=77), elbow dysplasia (n=44), hip dysplasia (n=39), unclear (n=14), medial patella luxation (n=13), shoulder pathology (n=10), angular limb deformity (n=4) and other (n=23, carpal osteoarthritis (OA), tarsal OA, elbow OA, elbow dysplasia and concurrent hip dysplasia, medial patella luxation and concurrent hip dysplasia, cranial cruciate ligament disease and concurrent hip dysplasia, caudal cruciate ligament rupture, avascular necrosis of the femoral head, immune-mediated polyarthritis, multipartite sesamoids). Surgical treatment was performed in 101 dogs and 116 dogs were treated conservatively. The median number of days of questionnaire completion was 70, 204 and 396 at T1, T2 and T3, respectively.

**LOAD Questionnaire**

The LOAD questionnaire was completed by the owners of 83% (n=180), 56% (n=121) and 36% (n=79) of the dogs at T1, T2 and T3, respectively.

***Numerical scores:*** There was a significant difference of the LOAD scores between T0 and all measured time points (*p* < 0.001) with a median absolute change of 5 (0 – 23), 6 (0 – 23) and 7 (0 – 31) at T1, T2 and T3 respectively. The recalled LOADscores were higher in 66% (n=115), 75% (n=86) and 73% (n=54) of cases at T1, T2 and T3 respectively (Figure 1). There was moderate agreement of the LOADscores between T0 and T1, moderate agreement between T0 and T2 and poor agreement between T0 and T3 (Table 1).

***Categorical scores:*** There was moderate agreement of the LOADcategorical scoring between T0 and T1, fair agreement between T0 and T2 and fair agreement between T0 and T3 (Table 2). The LOADcategorical scoring at T1 remained unchanged in 52% (n=89) of cases, changed by one category in 41% (n=71) of cases and by two categories in 7% (n=12) of cases (Supplementary file 1). At T2, the LOAD categorical scoringremained unchanged in 45% (n=52), changed by one category in 46% (n=53) of cases and by two categories in 6% (n=7) of cases (Supplementary file 1). At T3, the LOAD categorical scoring remained unchanged in 37.5% (n=27) of cases, changed by one category in 49% (n=35) of cases and changed by two categories in 14% (n=10) of cases (Supplementary file 1).

**CBPI Questionnaire**

The CBPI was completed by the owners of 83% (n=179), 56% (n=121) and 36% (n=77) of the dogs at T1, T2 and T3, respectively.

***PSS:*** There was a significant difference of the PSS numerical scores between T0 and all measured time points (*p* < 0.001) with a median absolute difference of 6 (0 – 27), 19 (0 -35) and 9 (0 – 24) at T1, T2 and T3 respectively. The recalled PSS scores were higher in 66% (n=118), 79% (n=95) and 77% (n=59) of the dogs at T1, T2 and T3 respectively (Figure 2). There was poor agreement of the PSS between T0 and all subsequent time points (Table 1).

***PIS:*** There was a significant difference of the PIS numerical scores between T0 and all measured time points (*p* < 0.001) with a median absolute difference of 7.5 (0 – 42), 35 (0 – 60) and 10 (0 – 45) at T1, T2 and T3 respectively. The recalled PIS scores were higher in 60% (n=107), 68% (n=82) and 68% (n=52) of dogs at T1, T2 and T3 respectively (Figure 3). There was moderate agreement of the PIS between T0 and T1, moderate agreement between T0 and T2 and poor agreement between T0 and T3 (Table 1).

***QOL:*** QOL score agreement was fair between T0 and all subsequent time points (Table 2). At T1, 66 owners (37%) recalled accurately their dog’s QOL; the score changed by one category for 41% of cases and by two categories for 18% (Supplementary file 2). At T2 and T3, the QOL of life remained the same for 44 (36%) and 27 (35%), respectively (Supplementary File 2).

**Factors impacting owners’ CMI score recollection**

Univariable multilevel logistic regression showed that the dog’s age, weight, gender, breed, affected limb, grade of lameness, diagnosis and type of management were not associated with difference between pre-treatment and subsequent LOAD scores. The only variable associated with the difference was the number of days between T0 and LOAD completion at follow-up (*p* = 0.0001), with weight (*p*=0.27) being the only other variable eligible for inclusion in the final multivariable model. Weight was excluded from the final multivariable logistic regression, leaving days since completion as the only variable showing a significant association.

**Discussion**

We reject our hypothesis on the basis that the agreement was not poor between T0 and all subsequent time points. However, our study shows that owners may not be able to accurately recall their dogs’ pre-treatment status using the LOAD and CBPI CMIs. The agreement between the actual and the recalled pre-treatment LOAD and CBPI numerical scores was poor to moderate with more than 60% of owners recalling their dog’s pre-treatment status as worse than they rated it at the time of initial presentation. More specifically, the agreement between the scores was moderate between T0 and T1 and T0 and T2 and poor between T0 and T3 for both questionnaires. Although the authors do not suggest the LOAD categorical scoring as the intended method of use of this CMI, for the purpose of this study, LOAD data was stratified into categorical data to determine if this could compensate for any recall bias that affected absolute numerical values. For example, the categorical scoring would remain the same (moderate) for numerical scores of 12/52 and 20/52 at T0 and T2 respectively and therefore could eliminate the difference in the numerical scores between these two time points and accomodate recall bias. The LOADcategorical scoring changed for approximately 50% of the cases at T1 and for >50% of the cases at T2 and T3. These results suggest that neither the numerical nor the categorical LOAD scores would be reliable when used retrospectively. Similarly, the CBPI categorical scoring only remained the same for just over one third of the study population at T1, T2 and T3.

Studies assessing the retrospective use of patient-reported outcome measures (PROMs) in people report similar findings.23-25, 27-30 Patients tended to recall their preoperative status as worse after lumbar spinal surgery, total knee replacement and THR. Similarly, patients who had cervical spinal surgery overestimated their preoperative disability following surgery, although they could accurately recall their preoperative pain and quality of life28. Higher recalled pre-treatment CMI scores could lead to overestimation of the effectiveness of an intervention particularly when there are no additional objective outcome measures.

Whilst it is easy to question the validity of this retrospectively acquired data it is important to consider why variations in these scores may have occurred. As the data requires participants to complete the questionnaire retrospectively, variation between contemporaneously and retrospectively acquired scores could be explained by memory distortion, cognitive dissonance reduction theory and/or response shift. 34, 35, 36 Biases leading to inaccurate recollection of pre-interventional state due to memory effects include recall bias, implicit theory of change and present state effect although it has been suggested that all three be considered under the umbrella term of recall bias.37 Implicit theory of change assumes that responders/observers cannot recall the preintervention status but instead reconstruct it based on the current state and how this has probably changed in the intervening period.35 Similarly, the present state effect suggests that responders use their current perceived health status to reconstruct their pre-interventional status.38 For example, an owner may assume their dog must have improved following an intervention and will therefore use implicit theory to conclude that their preintervention status was worse and therefore induce a negative response bias when completing the retrospective questionnaire.

Effort justification is derived from the cognitive dissonance theory.34 The effort justification theory suggests that when there is some form of personal sacrifice (mental, physical or monetary) in order to achieve a goal, dissonance is aroused; to reduce this dissonance a more positive response is seen towards the end goal.34 For example, an owner electing to have a THR performed on their dog undergoes an emotional, physical and in some scenarios monetary stress resulting in dissonance-arousal. To reduce this dissonance, the owner may want to perceive a better outcome to surgery than may have actually occurred, justifying this “sacrifice”. This again may result in a negative response bias when completing the questionnaire retrospectively.

The LOAD questionnaire has been used to retrospectively recall the preoperative status of patients undergoing THR >1 year after surgery15,19 and the LOAD and CBPI questionnaires to recall preoperative status following total elbow replacement (TER) at a median of 33 months postoperatively. Whilst the authors appreciate this is a simplified approach to a complex process and assumes that all score variation is due to recall bias, applying the results of our study to the findings of Henderson et al (2017) and De Sousa et al (2016) could illustrate a potential consequence of recalled CMI data. Applying the median LOAD score change at T3 to the data of Henderson et al (2017) is unlikely to affect the significance of that study. However, it is possible that contemporaneous pre-operative LOAD scores, had they been collected, would have been lower than those returned by the retrospective completion used at follow-up in that study. Similarly, whilst De Sousa et al (2016) did not demonstrate a significant difference between pre- and postoperative LOADNS it is feasible that, based on the data in this study, preoperative mobility status may have been significantly better than recalled indicating a potential negative impact of TER on mobility. They reported a significant improvement in PIS and PSS score. If the median PIS and PSS score change at T3 in was applied, these scores would not be significantly improved postoperatively and would question the benefit of TER in canine patients.

In contrast, response shift is not a consequence of memory effects but is instead defined as a change in the meaning of a target construct as a result of change in internal standards (recalibration), a change in values or priorities (reprioritisation) or a change in the definition (reconceptualization) of the target construct.39 Response shift is well recognised in longitudinal quality of life studies in people.35 Traditionally, treatment effect is calculated by subtracting the pre-test score from the post-test score. However, this doesn’t consider the effect of response shift.40 The “thentest” is the most common methodology used in human studies to quantify recalibration response shift (RRS).36, 40, 41 The thentest is similar to the methodology used in the present study as well as that of DeSousa et al (2016) and Henderson et al (2017). That is, respondents are asked to complete a questionnaire based on how they were prior to an intervention. 40, 41 RRS is calculated as the then-test minus the pre-test score and the adjusted treatment or time effect is calculated as the posttest minus thentest. 40, 41 The main perceived advantage of this approach is that it assumes the thentest and posttest share the same internal standards and are therefore a better estimate of treatment effect. 40, 41 For example, a client may initially rate a lameness as mild using the LOAD questionnaire. Following treatment, the client may see an improvement in lameness but not resolution and again rate the lameness as mild. On the retrospective thentest the client may instead rate the initial lameness as moderate due to a recalibration of internal standards. Using the traditional method there would be no treatment effect due to the response shift, whereas the thentest methodology would identify a positive treatment effect having accounted for response shift. One may therefore deduce from this that our study has simply calculated RRS and not the effect of recall bias. Similarly, one could conclude that DeSousa and Henderson calculated the adjusted treatment response, taking into account RRS, and as such may give a more accurate reflection of the magnitude and direction of treatment effect.

In reality, the retrospective nature of the thentest methodology makes it susceptible to recall bias and implicit theories of change as well as social desirability responding and effort justification. 42 RSS and recall bias should therefore be considered two distinct sources of variance in thentest scores. 41 It is likely that the retrospective CBPI and LOAD scores in this and other studies have been affected by recall bias and RRS. It has been suggested that determining the extent recall bias affects thentest scores is essential to determine the validity of this methodology in calculating RRS. 41 The current instrument designs means it is not possible to calculate to what extent recall bias affects LOAD and CBPI. The LOAD and CBPI questions are composed predominantly of perception- and evaluation-based questions and are therefore susceptible to recall bias, response bias and response shift. Further studies could focus on trying to assess the degree that recall bias affects retrospective LOAD and CBPI scores and determine to what extent variation in score reflects RRS. Recall bias could be considered a special form of error in measurement. 21, 22 Therefore including performance-based questions may allow calculation of variation secondary to recall bias.41 For example you could ask an owner to recall how long it took their dog to walk 100m; this value is unlikely to be affected by response bias or response shift and therefore failure to provide the correct answer would be considered recall bias. Alternatively, some authors suggest completing neuropsychological tests with assessment of response shift limited to those who demonstrate cognitive intactness, 41 although this may beyond the scope and abilities of many veterinary studies.

The time between T0 and subsequent time points of CMI completion was the only factor shown to significantly affect owners’ recollection in the present study, that is, the longer the time from initial completion of the CMIs, the greater was the difference between the recalled and T0 CMI scores. The agreement between the preoperative status and the recalled preoperative status was good to excellent at six and 12 weeks following THR in people. 43, 44 In this study, the agreement between the pre-treatment status and the recalled pre-treatment status of a dog was poor to moderate for dog owners at eight weeks from treatment; the agreement was moderate for the LOAD and CBPI PIS scores and poor for the CBPI PSS between T0 and T1. The ability to accurately recall a dog’s preoperative status within a time frame less than two months, as in humans, is unknown. It is possible that other factors may have affected owners’ recollection. In human medicine, there is conflicting evidence regarding the impact of age and mental health on patients’ recollection of their preoperative status. Recalled PROMs following arthroscopic rotator cuff repair were more likely to be accurate when reported by younger patients and a history of mental health condition was found to have no effect on recall accuracy in the same study.30 Comparable results regarding age were demonstrated by a study assessing patients’ preoperative recall after knee arthroplasty25; however, in this study patients with low mental health had poorer recall of function. Assessment of the effect of such owner-specific variables on owner recollection was beyond the feasibility of the current study.

Data in this study were collected mostly via a telephone interview including <50% of the recalled questionnaire scores at T1 and all the scores at T2 and T3 time points. To the authors’ knowledge, the effect of the method of questionnaire completion (via telephone interview, e-mail or by using paper-based questionnaires) on the patients’ or observers’ recollection has not been investigated and goes beyond the objective of this study. A recent study in people suggested that PROMs data were equally reliable when these were collected contemporaneously via a telephone interview or using paper-based questionnaires. 45 Paper-based questionnaires may be more susceptible to primacy effects (responders select the first agreeable option) whereas telephone interviews may be more vulnerable to recency effects (responders select the later options).46 If the extremes of these response order effects were applied to the LOAD questionnaire and our study design this could go some way to explain the difference in T0 and recalled LOAD scores reported. However, it has been shown that respondents were less likely to show recency effects when the response options were given at a slow pace over the phone.47 Questionnaire completion via e-mail may have eliminated this variable; however, it has been demonstrated that telephone surveys have a much higher response rate therefore increasing the volume of data that could be acquired over a given period of time and reducing the risk of nonresponse bias.48

We chose the LOAD and CBPI questionnaires to evaluate owners’ recollection in the present study as these are the CMIs used at our institution and these two CMIs have been used retrospectively in previous veterinary studies.15-19 The results are therefore specific to the LOAD and CBPI. Further investigation is required to assess owners’ recall accuracy using other CMIs. Despite that the LOAD and CBPI questionnnaires were validated using dogs that received medical management, in the present study they were used for cases that received surgical management as previously done in other studies, 15-17, 19 where the questionnaires were used to assess the outcome of surgery. Although one could question the use of LOAD and CBPI in cases with orthopaedic disease other than osteoarthritis, our study population was diagnosed with diseases that are associated with development of secondary osteoarthritis, a very common condition with prevalence ranging from 20% to 80%. 49, 50

The time points for the recalled CMI scores were determined based on standard clinical practice at our hospital. Most orthopaedic surgeons reassess the patients at approximately eight weeks following treatment; the medium and long-term outcome of a treatment is determined at approximately six months and a year following treatment in the veterinary orthopaedic literature.51

The questionnaires were completed by 80%, 57% and 37% of the study participants completed at T1, T2 and T3, respectively. Although one could consider that the results of the current study may have been affected by selection and non-response bias due the lower response rate at T2 and T3 with more than 50% participant ‘drop-out’ at T3,52 it has been shown by studies in people that surveys with low response rates may still be valuable and that surveys with high response rates may still be subject to bias. 53, 54

A final limitation of this study was that the study population was not standardised and affected by variable conditions. Future studies may wish to investigate the recall accuracy using CMIs for a specific orthopaedic condition.

In conclusion, owners’ perception of the preoperative status of their dog at two or more months after treatment may not represent the true preintervention status of their dog. Retrospective use of the LOAD and CBPI is subject to variation in preintervention scores and until there is a better understanding to what extent this variation can be attributed to the effects of recall bias, response shift and implicit theory of change, the authors can only recommend the use of contemporaneously collected CMI data. Retrospectively collected CMI data may not be reliable even when collected less than six months from the time of treatment.

As such, studies which use retrospectively collected CMI data obtained at two or more months after treatment should be interpreted with caution.

**References**

1 Brown DC, Boston RC, Coyne JC, Farrar JT. Ability of the Canine Brief Pain Inventory to

detect response to treatment in dogs with osteoarthritis. J Am Vet Med Assoc 2008;

233(8): 1278-1283

2 Brown DC, Bell M, Rhodes L. Power of treatment success definitions when the Canine Brief

Pain Inventory is used to evaluate carprofen treatment for the control of pain and

inflammation in dogs with osteoarthritis. Am J vet Res 2013; 74: 1467-1473

3 Innes JF, Barr ARS. Can owners assess outcome following treatment of canine cruciate

ligament deficiency? Journal of Small Animal Practice 1998; 39:373-378

4 Brown DC, Boston RC, Farrar JT. Comparison of force plate gait analysis and owner

assessment of pain using the Canine Brief Pain Inventory in dogs with osteoarthritis. J vet

Intern Med 2013; 27(01): 22-30

5 Quinn MM, Keuler NS, Lu Y, Faria MLE, Muir P, Markel MD. Evaluation of agreement

between numerical rating scales, visual analogue scoring scales, and force plate gait

analysis in dogs. Vet surg 2007; 36: 360-367

6 Waxman AS, Robinson DA, Evans RB, Hulse DA, Innes JF, Conzemius MG. Relationship

between objective and subjective assessment of limb function in normal dogs with an

experimentally induced lameness. Vet Surg 2008; 37: 241-246

7 Oosterlinck M, Bosmans T, Gasthuys F et al. Accuracy of pressure plate kinetic asymmetry

indices and their correlation with visual gait assessment scores in lame and nonlame dogs.

Am J Vet Res 2011; 72: 820-825.

8 Hercock C, Pinchbeck G, Giejda A, clegg PD, Innes JF. Validation of a client-based clinical

metrology instrument for the evaluation of canine elbow osteoarthritis. Journal of Small

Animal Practice 2009; 50: 266-271

9 Hudson JT, Slater MR, Taylor L, Scott HM, Kerwin SC. Assessing repeatability and validity of

a visual analogue scale questionnaire for use in assessing pain and lameness in dogs. Am J

Vet Res 2004; 65: 1634-1643

10 Hielm-Bjorkman AK, Rita H, Tulamo RM. Psychometric testing of the Helsinki chronic pain

index by completion of a questionnaire in Finnish by owners of dogs with signs of pain

caused by osteoarthritis. Am J Vet Res 2009; 70: 727-734

11 Walton MB, Cowderoy E, Lascelles D, Innes JF. 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 1-10

12 Pinna S, Lambertini C, Grassato L, Romagnoli N. Evidence-based veterinary medicine: A

tool for evaluating the healing process after surgical treatment for cranial cruciate

ligament rupture in dogs. Front Vet Sci 2019; 6: 65

13

14 Brown DC, Boston RC, Coyne JC, Farrrar JT. Development and psychometric testing of an

instrument designed to measure chronic pain in dogs with osteoarthritis. Am J Res 2007;

68: 631-637

15 Forster KE, Wills A, Torrington AM et al. Complications and owner assessment of canine

total hip replacement: a multicentre internet based survey. Vet Surg 2012; 41: 545-550

16 DeSousa RJR, Parsons KJ, Owen MR, et al. Radiographic, surgeon and owner assessment

of the Biomedtrix TATE elbow arthroplasty. Vet Surg 2016; 45: 726-735

17 Swiderski, Palmer RH. Long-term outcome of distal femoral osteotomy for treatment of

combined distal femoral varus and medial patellar luxations: 12 cases. J Am Vet Med

Assoc 2007; 231: 1070-1075

18 Terhaar HM, Muir P, Baker LA, Binversie EE, Chi J, Sample SJ. Contribution of habitual

activity to cruciate ligament rupture in Labrador Retrievers. Vet Comp Orthop

Traumatol 2020; 33: 82-88

19 Henderson ER, Wills A, Torringtton AM et al. Evaluation of variables influencing success

and complication rates in canine total hip replacement: results from the British Veterinary

Orthopaedic Association Canine Hip Registry (collation of data: 2010-2012). Vet Rec 2017;

181: 18

20 Godlonton S, Hernandez MA, Murphy M. Anchoring bias in recall data: Evidence from

central America. Amer. J. Agr. Econ. 100(2): 479-501

21 Krogsgaard MR, Brodersen J, Jensen J, Hansen CF, Cominns J. Potential problems in the

use of patient reported outcome measures (PROMs) and reporting of PROM data in

sports science. Scand J Med Sci Sports 2021; 31: 1249-1258

22 Coughlin SS. Recall Bias in epidemiologic Studies. J Clin Epidemiol 1990; 43: 87-91

23 Pellise F, Vidal X, Hernandez A, cedraschi C, Bago J, Villanueva C. Reliability of

retrospective clinical data to evaluate the effectiveness of lumbar fusion in chronic low

back pain. Spine 2005; 30; 3; 3: 365-368

24 Murphy MT, Vardi R, Journeaux SF, Whitehouse SL. A patient’s recollection of pre-

operative status is not accurate one year after arthroplasty of the hip or knee. Bone Joint

J 2015; 97-B: 1070-5

25 Lingard EA, Wright EA, Sledge CB, Kinemax Outcomes Group. Pitfalls of using patient

recall to derive preoperative status in outcome studies of total knee arthroplasty. The

Journal of Bone & Joint surgery 2011: 83; 8: 1149-1156

26 https://www.pennvet.com/customer/wcm/connect/0caa373a-7925-4c0a-826f-1c34f6532376/LOAD\_PrintableForm.pdf?MOD=AJPERES&CACHEID=ROOTWORKSPACE-0caa373a-7925-4c0a-826f-1c34f6532376-mN3Pgo.

27 Rodrigues R, Silva PS, Cunha M, Vaz R, Pereira P. Can we assess the success of surgery for

degenerative spinal diseases using patients’recall of their preoperative status? World

NeuroSurg 2018 115:e768-773

28 Aleem IS, Duncan J, Ahmed AM et al. Do cervical spine surgery patients recall

their preoperative status? A cohort study of Recall Bias in Patient-reported Outcomes.

Clin Spine Surg 2018; 31: E481-487

29 Aleem IS, Duncan J, Ahmed AM et al. Do lumbar decompression and fusion patients recall

their preoperative status? A cohort study of Recall Bias in Patient-reported Outcomes.

Spine 2017; 42: 2: 128-134

30 Gotlin MJ, Kingery MT, Baroon SL et al. Recall bias in retrospective assessment of

preoperative patient-reported American shoulder and elbow surgeons scores in

arthroscopic rotator cuff repair surgery. The American Journal of sports Medicine 2020;

48(6): 1471-1475

31 Sumner-Smith, G: Gait Analysis and Orthopaedic Examination. In Textbook of small Animal

Surgery 2nd ed. 1993, p.1578

32 Landis JR, Koch GG. The measurement of observer agreement for categorical data.

Biometrics 1977; 33(1): 159-174

33 Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for

reliability research. J Chiropr Med 2016; 15(2): 155-163

34 Harmon-Jones E, Mills J. An introduction to Cognitive Dissonance Theory and an Overview

of Current Perspectives on the Theory. In: Harmon-Jones E ed. Cognitive Dissonance:

Progress on a Pivotal Theory in Social Psychology. Washington DC: American Psychological

Asscociation; 2019, 3-24.

35 Schwartz CE, Bode R, Repucci N, Becker J, Sprangers MAG, Fayers PM. The clinical

significance of adaptation to changing health: A meta-analysis of response shift. Quality of

Life Research 2006; 15: 1533-1550

36 Schwartz CE, Sprangers MAG, Carey A, Reed G. Exploring response shift in longitudinal

data. Pathology and Health 2004; 19: 51-69.

37 Blome C, Augustin M. Measuring Change in Quality of Life: Bias in Prospective and

Retrospective Evaluation. Value in Health 2018; 18: 110-115.

38 Meyer T, Richter S, Raspe H. Agreement between pre-post measures of change and

transition ratings as well as then-tets. BMC Medical Research Methodology 2013; 13: 52.

39 Sprangers MAG, Schwartz CE. Integrating response shift into health-related quality of life

research: a theoretical model. Social Science and Medicine 1999; 48: 1507-1515.

40 Sprangers MA, Van Dam FS, Broersen J, LodderL, Wever L, Visser MR. Revealing response

shift in longitudinal research on the fatigue-the use of the thentest approach. Acta

Oncologica 1999; 38(6): 709-718.

41 Schwartz CE, Sprangers MAG. Guidelines for improving the stringency of response shift

research using the thentest. Qual Life Res 2010; 19: 455-464.

42 Normal G. Hi! How are you? Response shift, implicit theories and differing epistemologies.

Qual of Life Res 2002; 12: 239-249.

43 Marsh J, Bryant D, MacDonald SJ. Older patients can accurately recall their preoperative

health status six weeks following total hip arthroplasty. J Bone Joint Surg Am. 2009; 91:

2827-2837

44 Howell J, Xu M, Duncan CP, Masri BA, Garbuz DS. A comparison between patient recall and

concurrent measurement of preoperative quality of life outcome in total hip arthroplasty.

J Arthroplasty 2008; 23: 843-849

45 Adoogwa O, Elsamadicy AA, Cheng J, Bagley C. Assessinng patient reported outcomes

measures via phone interviews versus patient self-survey in the clinic: Are we measuring

the same thing? World Neurosurg 2016; 87: 230-234.

46 Krosnick JA, Alwin DF An evaluation of a cognitive theory of response-order effects in

survey measurement. Public Opinion Quarterly 1987; 51: 201-219

47 Mingay DJ, Greenwell MT Memory Bias and Response-order Effects. Journal of Official

Statistics 1989; 5: 253-263

48 Sinclair M, O’Toole J, Malawaraarachchi M, Leder K. Comparison of response rates and

cost-effectiveness for a community-based survey: postal, internet and telephone modes

with generic or personalised recruitment approaches. BMC Medical Research

Methodology 2012, 12:132.

49 Pettitt RA, German AJ. Investigation and management of canine osteoarthritis. In Practice

2015; 37: 1-8.

50 Johnston SA. Osteoarthritis – joint anatomy, physiology and pathobiology. Veterinary

Clinics of North America-Small Animal Practice. 1997; 27: 699-723.

51 Cook JL, Evans Conzemius MG et al Proposed definitions and criteria for reporting time

frame, outcome, and complications for clinical orthopaedic studies in veterinary medicine.

Vet Surg 2010; 39: (8): 905-8

52 Compton J, Glass N, Fowler T. Evidence of selection bias and non-response bias in patient

satisfaction surveys. Iowa Orthop J 2019; 39(1): 195-201.

53 Lie HC, Rueegg CS, Fossa SD, Loge JH, Ruud E, Kiserud CE. Limited evidence of non-response

Bias despite modest response rate in a nationwide survey of long-term cancer survivors –

Results from the NOR-CAYACS study. Journal of Cancer Survivorship 2019; 13: 353-363.

54 Gustavson K, Roysamb E, Borrenn I. Preventing bias from selective non-response in

population-based survey studies: findings from a Monte Carlo simulation study. BMC

Medical Research Methodology 2019; 19: 120.

Figure 1. LOAD score change between T0 and T1, **a**; T0 and T2, **b**; T0 and T3, **c**. T0, initial presentation; T1, 2 months from initial presentation; T2, 6 months from initial presentation; T3, 12 months from initial presentation.

Figure 2. CBPI PSS score change between T0 and T1, **a**; T0 and T2, **b**; T0 and T3, **c**. PSS, Pain Severity Score; T0, initial presentation; T1, 2 months from initial presentation; T2, 6 months from initial presentation; T3, 12 months from initial presentation.

Figure 3. CBPI PIS score change between T0 and T1, **a**; T0 and T2, **b**; T0 and T3, **c**. PIS, Pain Interference Score; T0, initial presentation; T1, 2 months from initial presentation; T2, 6 months from initial presentation; T3, 12 months from initial presentation.

Table 1. Agreement of LOAD and CBPI PSS and PIS scores between T0-T1, T0-T2 and T0-T3.

|  |  |  |  |
| --- | --- | --- | --- |
|  | T0-T1 | T0-T2 | T0-T3 |
| LOAD ICC; 95% CI | 0.640  (0.483-0.747) | 0.533  (0.278-0.696) | 0.496  (0.222-0.680) |
| CBPI PSS ICC; 95% CI | 0.450  (0.286-0.581) | 0.366  (0.107-0.558) | 0.419  (0.084-0.641) |
| CBPI PIS ICC; 95% CI | 0.569  (0.459-0.661) | 0.559  (0.416-0.673) | 0.432  (0.219-0.603) |

Abbreviations: PSS, Pain Severity Score; PIS, Pain Interference Score; T0, initial presentation; T1, 2 months from initial presentation; T2, 6 months from initial presentation; T3, 12 months from initial presentation; ICC, Intraclass Correlation Coefficient; CI, Confidence Interval.

Table 2. Agreement of LOADcategorical scoring and CBPI QOL scores between T0-T1, T0-T2 and T0-T3.

|  |  |  |  |
| --- | --- | --- | --- |
|  | T0-T1 | T0-T2 | T0-T3 |
| LOAD κw; 95% CI | 0.449  (0.355-0.544) | 0.359  (0.239-0.480) | 0.295  (0.156-0.434) |
| CBPI QOL κw; 95% CI | 0.319  (0.225-0.412) | 0.348  (0.239-0.456) | 0.337  (0.199-0.474) |

Abbreviations: QOL, quality of life; T0, initial presentation; T1, 2 months from initial presentation; T2, 6 months from initial presentation; T3, 12 months from initial presentation; κw, kappa weighted; CI, Confidence Interval.