Evaluation of Clinical Metrology Instrument in Dogs with Osteoarthritis

C. Muller, B. Gaines, M. Gruen, B. Case, K. Arrufat, J. Innes, and B.D.X. Lascelles

Background: In veterinary clinical pain studies, there is a paucity of data on test-retest variability in Clinical Metrology Instruments (CMIs), and it is unknown whether CMIs should be administered using independent (respondents not permitted to see previous answers) or dependent (respondents shown previous answers) interviewing.

Objectives: To compare baseline variability in CMIs designed to assess pain in dogs with osteoarthritis, and compare CMI scores using independent (InD) and dependent interviewing (DI) for the Canine Brief Pain Inventory (CBPI) and the Client-Specific Outcome Measures (CSOM).

Animals: Fifty-one client-owned dogs with radiographic evidence of osteoarthritis and associated pain.

Methods: Clinical Metrology Instruments data were collected during 2 randomized, double-masked, placebo-controlled, proof of principle pilot studies with parallel treatment groups. Enrolled dogs received either placebo or antinerve growth factor antibody (NV-01).

Results: Agreement between baseline CMI scores was good (CBPI Pain $P = .29$, CBPI Interference $P = .32$, CSOM $P = .036$, LOAD $P = .67$, HCPI $P = .27$), being best for the LOAD (ICC = 0.89). CMI responses collected during independent and dependent interviewing were not statistically different (CBPI Pain $P = .33$, CBPI Interference $P = .28$, CSOM $P = .42$) and showed good agreement. Additionally, dependent interviewing resulted in increased treatment effect sizes.

Conclusions and Clinical Importance: There is little difference between independent and dependent interviewing, however, dependent interviewing resulted in increased treatment effect sizes. By using dependent interviewing, investigators could increase clinical trial power through minimal change to study design. Further research is warranted to investigate the use of dependent interviewing.

Key words: Analgesia; Client metrology instrument; Degenerative joint disease; Questionnaire.

In veterinary medicine, performing clinical trials that use subjective measures provide a unique challenge, in that completion of these measures requires a proxy assessment of the animal. In many cases, this is done using a clinical metrology instrument (CMI), also referred to as a questionnaire. The use of a proxy to determine the effect of an intervention (be it a drug or surgical method) is complicated by many factors that can introduce bias, and also the inherent variability from one pet owner to another.[1–5] Therefore, using valid questionnaires and administering them in an appropriate manner is essential in decreasing variability[6]

Abbreviations:

AlkPhos	alkaline phosphatase
ALT	alanine transferase
BCS	body condition score
BW	body weight
CBC	complete blood count
CBPI	canine brief pain inventory index
CI	confidence interval
CMI	clinical metrology instrument
CP	conscious proprioception
CPRL	comparative pain research laboratory
CSOM	client-specific outcome measure
CVM	College of Veterinary Medicine
DI	dependent interviewing
DJD	degenerative joint disease
ES	effect size
HCPI	Helsinki Chronic Pain Index
IACUC	Institutional Animal Care and Use Committee
ICC	intraclass correlation coefficient
InD	independent interviewing
LOAD	Liverpool Osteoarthritis in Dogs Index
NCSU	North Carolina State University
NGF	nerve growth factor
NSAIDs	nonsteroidal anti-inflammatories
OA	osteoarthritis
QoL	quality of life
SD	standard of deviation
SIPP	survey income and program participation

and confounding factors,[1,7] and providing accurate results.

A large body of work has been performed evaluating the psychology behind how a respondent answers a questionnaire. Overall, this work demonstrates that

small, seemingly innocuous changes to a questionnaire or the manner in which it is presented can drastically change the results.[8] Factors such as how the questionnaire is presented,[2,9] how the questionnaire is written[1,10], and even what type of paper the questionnaire is printed on[7] have all been determined to have an effect.

Bias has been defined as a deviation of results or inferences from the truth.[11] One form of bias is the faking bad bias (also known as the hello-goodbye effect), where respondents try to appear sick to qualify for support.[3,12] In reference to pain studies within veterinary medicine, respondents could indicate higher impairment scores so as to gain entry for their pet into the clinical study. Once on the study, respondents often then revert to a "more true" representation of their pet's impairment. Additionally, dogs may tend to be recruited to a clinical study when the clinical signs are obvious, and natural variation in the disease causes them to become less affected as the study starts. These 2 phenomena are likely partly responsible for the placebo effect. However, most studies have shown good repeatability (test-retest) for the validated CMIs used to assess pain and mobility.[13–15]

Another form of bias seen specifically in independent interviewing is called the seam bias.[16] Seam bias is defined as an overrepresentation of change, when measured across a "seam" between 2 successive survey administrations and is a unique bias seen in independent interviewing (InD) during longitudinal panel surveys.[17,18] Subsequent work showed that dependent interviewing (DI), allowing respondents to review their previous answers, was more effective in controlling seam bias and therefore resulted in a more accurate collection of data.[16,18]

In most instances, veterinary clinical studies are designed to not allow respondents to view their previous answers, a technique termed "independent interviewing."[19] The argument for this technique has traditionally been that it provides a less biased view of the current status of their pet, and more accurate assessments will be obtained. However, by showing respondents their original answers during the retest phase, or using "dependent interviewing," one might receive a more accurate representation in the linear progression[18,20] of change in a animal's status.

For our study, we hypothesized that there would be a significant difference in CMI responses at the screening visit versus the start of study visit for CMIs designed to assess pain and mobility in dogs with osteoarthritis. Additionally, we hypothesized that there would be a significant difference in responses under conditions of dependent interviewing, compared to independent interviewing, when assessing dogs with osteoarthritis during a clinical study. Further, we hypothesized that dependent interviewing would lessen the placebo effect, and increase treatment effects.

The aims of this study were:

1 To compare different CMI scores at screening (Day −7) and Day 0 (baseline measures).

2 To compare CMI scores using independent and dependent interviewing techniques when using the Canine Brief Pain Inventory (CBPI) and the Client-Specific Outcome Measures (CSOM).

3 To compare the influence of dependent interviewing on scores in the placebo versus active treatment groups.

Methods and Materials

Data were included from dogs participating in 2 studies conducted at the North Carolina State University College of Veterinary Medicine (NCSU-CVM) investigating the use of a novel therapeutic for the treatment of osteoarthritis-associated pain. Both studies (A[21] and B) were randomized, double masked, placebo-controlled proof of principle pilot studies with parallel treatment groups. The Institutional Animal Care and Use Committee (IACUC) approved both studies (IACUC #12-149-O), and in all cases owners signed a written consent form following a detailed verbal explanation of the study protocol. Studies were conducted between May 2013 and November 2014.

Study Population

Data were prospectively collected during 2 separate studies evaluating 2 different populations of dogs, study A and study B. Evaluation of independent versus dependent interviewing techniques was done using data collected from study A and evaluation of baseline stability was done using data collected from both study A and study B.

In study A, as previously reported,[21] following 81 initial enquires, 37 dogs were screened, and 26 entered the study with 25 completing the 28-day study. One dog was withdrawn after the Day 14 assessment, because of a cruciate ligament rupture. The flow of dogs through the study is shown in Figure 1. Data from study A were used for evaluation of baseline stability, as well as to assess independent versus dependent interviewing techniques.

In study B, after 101 initial enquires, 36 dogs were screened, and 26 entered the study with all 26 dogs completing the screening (Day −7) and the study Day 0 assessments. Data from study B were used in the assessment of baseline stability, together with data from study A. The flow of dogs through study B is shown in Figure 2. No independent/dependent interviewing data were available from study B.

Dogs in both studies were randomized to receive the drug or placebo on Day 0 using a computer-derived[a] randomization schedule. The randomization schedule and the key were held by the NCSU pharmacy and not disclosed to investigators until completion of the statistical analysis.

Inclusion Criteria

To be eligible for either study, dogs were required to be greater than 1 year old, ≥15 kg, have owner-rated mobility impairment, and at least one appendicular joint or axial skeleton area that was considered painful on orthopedic exam and where radiographs showed the presence of osteoarthritis (OA), as previously described in detail.[21] Dogs were required to be in generally good health and not be currently receiving any anti-inflammatory medications (NSAIDs), if the dogs were currently on NSAIDs, a 2-week withdrawal period was required before study entry. Other analgesics (eg, amantadine, gabapentin, tramadol) were permitted only if pain was still present, the dog had been on the medication(s) for at least 3 weeks and the treatment regimen could be continued throughout the trial, otherwise a 2-week withdrawal period was
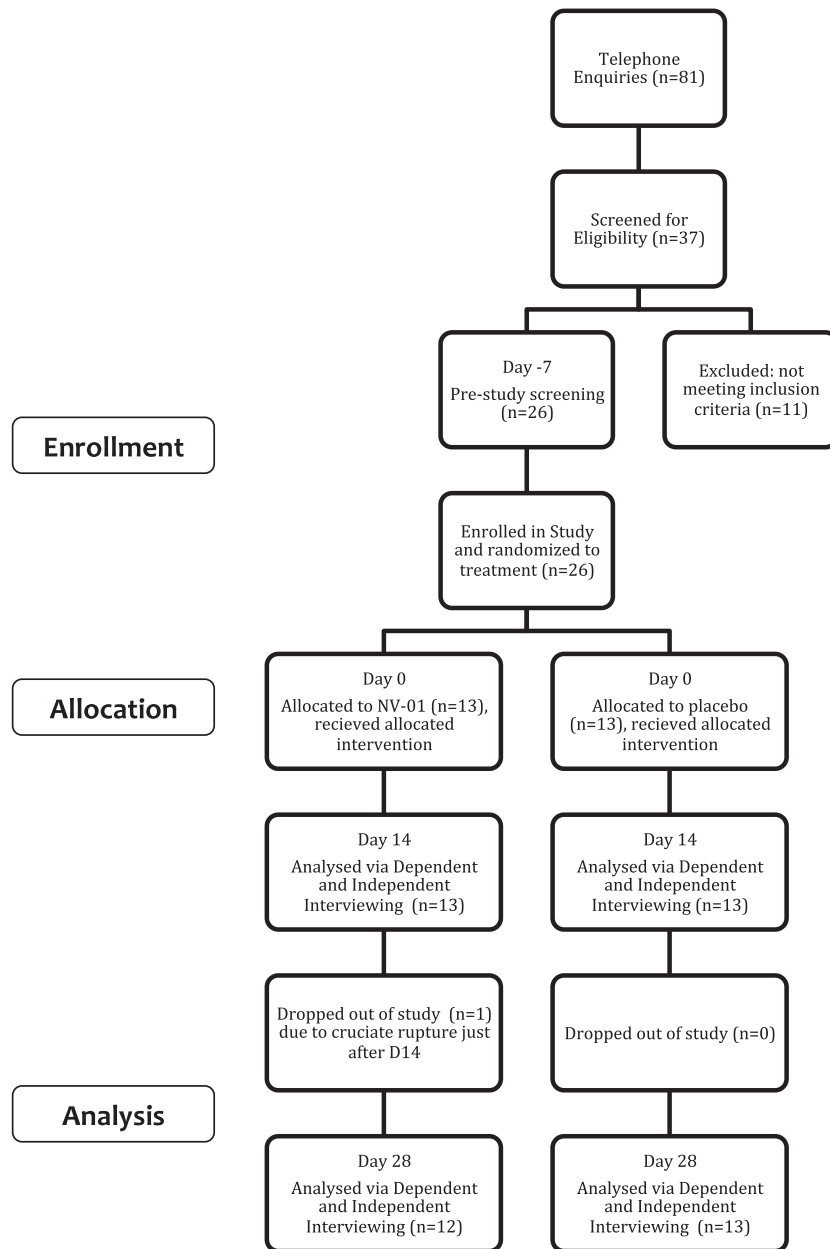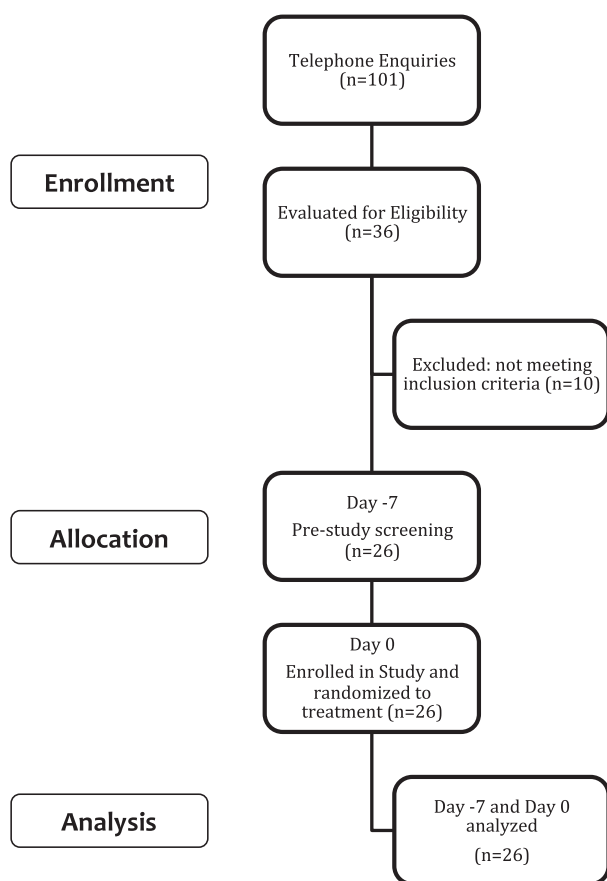
Fig 1. Study A Dog Selection and Enrollment Algorithm. Enrolled dogs were enrolled in either the treatment group with antinerve growth factor antibody (NV-01) or the placebo group. Data were collected to compare baseline variability, dependent, and independent interviewing techniques.

required before study entry. Dogs were required to be either not receiving nutritional supplements, or have been on them for at least 6 weeks before the start of the study and continue them throughout the study. If dogs were considered to be mobility impaired, but no OA was detected radiographically, or if they had OA but the impairment in mobility was not sufficient, they were not enrolled. Other exclusion criteria included known or suspected presence of any of the following conditions: clinically significant cardiovascular disease; severe dental disease; neurological disease, renal disease; liver disease (ALT levels of up to twice the upper normal value and AlkPhos levels of up to 4 times the upper normal value were considered acceptable in the absence of other signs of liver disease); chronic pulmonary disease; infectious disease; immune-mediated disease; neoplasia; urinary tract infection;

hypothyroidism (unless well controlled); diabetes mellitus; skin disease of the foot; obesity (8 or 9 of the 1–9 Body Condition Score Scale[22]). Particular attention was given to ruling out neurological disease through a comprehensive neurological evaluation. Additionally, owners had to agree to not change the management of dogs for the period of the study, and owners were required to have a stable lifestyle for the duration of the study (eg, no planned house moves, vacations, relationship changes, or new pets).

Study Protocol

Study A was conducted over a 35-day period with outcome measures gathered on Day −7, Day 0, Day 14 and Day 28. On Day −7 and Day 0 (baseline period) the owners completed 4

Fig 2. Study B Dog Selection and Enrollment Algorithm. Data were collected only to compare baseline variability.

CMI's (Canine Brief Pain Inventory [CBPI], Liverpool Osteoarthritis in Dogs Index [LOAD], Helsinki Chronic Pain Index [HCPI] and Client-Specific Outcome Measures [CSOM]) using only independent interviewing techniques, to establish a baseline for each dog. On Day 0, dogs were randomized to receive either the antinerve growth factor antibody (NV-01), or placebo. Thirteen dogs were administered NV-01 at a dose of 200 μg/kg of a 2 mg/mL solution IV over a 1-min period through an intravenous 20-gage catheter.

Thirteen dogs were administered a placebo (normal saline), administered IV at a volume equivalent to the dose of NV-01. The dispensing of drug or placebo was performed by NCSU pharmacy, with all other personnel involved in the collection of data masked to the administration until completion of the statistical analysis. Pharmacy personnel prepared unmarked syringes for each dog with the barrel covered in opaque tape. Testing before starting the study indicated there was no appreciable difference between the feel of injecting saline versus NV-01 through a 20-gage catheter, to ensure complete masking.

At Day 14 and Day 28 owners also completed all CMIs (independent interviewing). Once completed, owners were represented with the CSOM and the CBPI CMIs, in sequence, and for each one, they were shown the answers from the previous visit and asked to complete these CMIs again (dependent interviewing).

Study B had a very similar initial design, with screening being performed 7 days before study entry. In this study, baseline CMI data for the CSOM, CBPI, LOAD and HCPI were collected at Day −7 and Day 0. The owners completed these 4 CMI's using only independent interviewing techniques. Treatments were again

administered on Day 0, using the same drug and placebo, but they were administered subcutaneously.

Clinical Metrology Instruments

All CMI's were completed by the same owner at all visits, and owners were directed to base their answers on their observations of the preceding 7 days. The owners completed the CMI's while sitting or standing (as they preferred), in a standard consulting room. All CMI's were printed using standardized paper.[b] At each time point, they were presented with each CMI sequentially, and requested to complete the CMI with basic instructions explained to them in a neutral tone of voice using a standard script. Once completed, the next CMI was handed to them. CMI's were always presented in the following order: CBPI; LOAD; HCPI; CSOM. Owners completed the CMI's while their dog was being examined in a separate room. Only complete CMIs were considered valid.

The CBPI[14,23] is a 2-part instrument assessing pain severity and pain interference. The pain severity score (CBPI pain) is the arithmetic mean of 4 items scored on an 11-point (0–10) numerical scale with higher scores correlating to increased level of pain, and the pain interference score (CBPI interference) is the mean of 6 items similarly scored, with higher scores indicating increased pain leading to greater interference with daily activity.

The LOAD[13,24] is a 13-item instrument with all items reported on a 5-point Likert-type scale. Each item is scored 0–4, and the item scores are summed to give an overall instrument score, with increased scores indicating increased abnormalities seen in the dog's behavior, mobility, and exercise levels.

The HCPI[25] is a CMI with a total of 11 questions regarding activity, behavior, and mood. Each question was evaluated using a Likert-type scale. Each item was scored 0–4, with higher scores indicating increased pain and was summed to give an overall instrument score.

The CSOM[15] is a CMI that follows 3 activities that are determined to be impaired and are unique to the individual dog. It is modeled after the Cincinnati Orthopedic Disability Index (CODI).[26] The CSOM was constructed by a single study investigator (BC) for each case as previously described.[15] At each time point, the difficulty performing each of the 3 activities was scored on a 0–4 scale: 0 = No Problem, 1 = Mildly Problematic, 2 = Moderately Problematic, 3 = Severely Problematic, and 4 = Impossible. The total CSOM score represented the sum of scores for individual activities.

Statistical Analysis

Data collected from both studies were powered in a similar manner, as pilot studies, as previously described.[21] Two different approaches were utilized in comparing the agreement of the CMI scores between the screening (Day −7) and Day 0 from studies A and B combined. First, average scores were compared to investigate the presence of a systematic bias in the answers, by plotting mean scores for each time point (combining both treatment and placebo groups) ± 1 standard error and also more formally using the Wilcoxon signed-rank tests.[c] Next, the extent of agreement between screening (Day −7) and Day 0 from individual responses were evaluated using agreement plots and intraclass correlation coefficients (ICCs). Agreement plots included the individual values, the regression line and also the 45-degree lines. Statistically, the agreement was summarized using the ICCs.

For the second aim of this study, dependent interviewing and independent interviewing CSOM, CBPI pain, and CBPI interference scores were compared in a similar fashion to the baseline scores. The CSOM, CBPI pain, and CBPI interference scores for

Day 14 and Day 28, were evaluated by plotting mean scores for each time point (combining both treatment and placebo groups) ± 1 standard error. To formally test if the mean scores were significantly different, the Wilcoxon signed-rank test was used. Agreement plots were created and ICCs calculated.

For the third aim of the study, treatment effects were analyzed using the Wilcoxon rank-sum test to determine if there was a significant difference between the treatment and placebo groups, using either the dependent or independent interviewing techniques. In addition, standardized effect sizes were calculated to compare the results from the dependent and independent interviewing techniques.

Results

Variability in Baseline Data Across Different CMIs

The first portion of the study involved analyzing the baseline data between the screening (Day −7) and Day 0. For CBPI pain, and HCPI, the average scores for Day 0 were higher (greater impairment) than the Day −7 average. The LOAD showed a smaller increase from Day −7 to Day 0. The CBPI interference and CSOM total showed an average decrease (less impairment) between screening (Day −7) and Day 0. There was a significant difference between the average scores on Day −7 and Day 0 for CSOM Total ($P = .036$), but not for the CBPI measures (CBPI Pain $P = .29$, CBPI Interference $P = .32$), LOAD ($P = .67$), or HCPI ($P = .27$).

Graphical evaluation of the agreement for CBPI pain, CBPI interference, LOAD, and HCPI indicated visually good agreement, given that the regression and 45-degree lines intersect, and the 45-degree lines overlap with the 95% confidence limits for the regression lines (Fig 3). For CSOM, the regression line lies entirely below the 45-degree line, suggesting that the Day −7 values tend to be higher (greater impairment) than the corresponding Day 0 values. However, the 45-degree line for CSOM does barely intersect with the 95% confidence limits for the regression line, indicating that the difference is not statistically significant. The best visual agreement is for the LOAD.

The estimates and corresponding 95% confidence intervals for the intraclass correlation coefficients were then evaluated. These showed strong agreement for CBPI pain (ICC = 0.79, 95% confidence 0.64–0.88), CBPI interference (ICC = 0.76, 95% confidence 0.59–0.87), and CSOM (ICC = 0.76, 95% confidence 0.59–0.86), and almost perfect agreement for HCPI (ICC = 0.83, 95% confidence 0.70–0.90) and LOAD (ICC = 0.89, 95% confidence 0.81–0.94), with the best agreement being seen for the LOAD.

Influence of Independent and Dependent Interviewing Techniques on CMI Scores

For the CBPI pain (Day 14 $P = .33$; Day 28 $P = .09$), CBPI interference (Day 14 $P = .28$; Day 28 $P = .08$), and CSOM Total (Day 14 $P = .42$; Day 28 $P = 1.0$), there were no significant differences in the average values between the independent and dependent answers.

Agreement of Independent and Dependent CMI Scores

Graphical evaluation of the agreement between independent and dependent interviewing CMI scores showed good agreement at both Day 14 and Day 28 across the CBPI pain (Fig 4), CBPI interference (Fig 5) and CSOM (Fig 6). Additionally the ICC estimates, as well as the 95% confidence intervals, showed almost perfect agreement for all CMIs; CBPI Pain (Day 14 ICC = 0.98, 95% confidence 0.95–0.99; Day 28 ICC = 0.98, 95% confidence 0.95–0.99), CBPI Interference (Day 14 ICC = 0.97, 95% confidence 0.93–0.99; Day 28 ICC = 0.97, 95% confidence 0.94–0.99), and CSOM (Day 14 ICC = 0.91, 95% confidence 0.79–0.96; Day 28 ICC = 0.92, 95% confidence 0.81–0.97).

### Evaluation of Treatment Effects Using Dependent and Independent Interviewing

The placebo and treatment responses for both the dependent and independent interviewing technique were evaluated. Data were graphed, and responses from the surveys completed with dependent interviewing showed greater improvement (a larger reduction in the average CMI scores) compared to those completed with independent interviewing techniques. This difference appeared more pronounced for the treatment group than the placebo group (Fig 7) in all 3 CMIs.

When the change in CMI scores between Day 0 and Day 14, and Day 0 and Day 28 was compared between the treatment and placebo group, the significance of the treatment effect did not change, but P-values were smaller, CBPI Pain (Day 0–14 InD $P = .23$ and DI $P = .16$; Day 0–28 InD $P = .21$ and DI $P = .11$), CBPI Interference (Day 0–14 InD $P = .21$ and DI $P = .11$; Day 0–28 InD $P = .30$ and DI $P = .20$), and CSOM (Day 0–14 InD $P = .04$ and DI $P = .03$; Day 0–28 InD $P = .009$ and DI $P = .009$).

Finally, the between group differences for the independent and dependent interviewing techniques were compared by looking at the standardized effect sizes, CBPI Pain (Day 0–14 InD 0.57 and DI 0.71; Day 0–28 InD 0.46 and DI 0.65), CBPI Interference (Day 0–14 InD 0.72 and DI 0.87; Day 0–28 InD 0.40 and DI 0.54), and CSOM (Day 0–14 InD 0.86 and DI 0.91; Day 0–28 InD 1.2 and DI 1.2). In general, the treatment effect sizes were larger for the dependent interviewing technique, except for the CSOM Total at Day 28.

Discussion

We identified baseline variability in CSOM, CBPI, LOAD, and HCPI scores when measured 7 days apart with no intervening treatment, however, the scores at the 2 baseline time points were not statistically significantly different, rejecting our hypothesis. The direction of shift in baseline scores was not consistent across all the CMIs, with the best agreement in baseline scores being found for the LOAD CMI. Overall, we found good agreement between dependent interviewing and
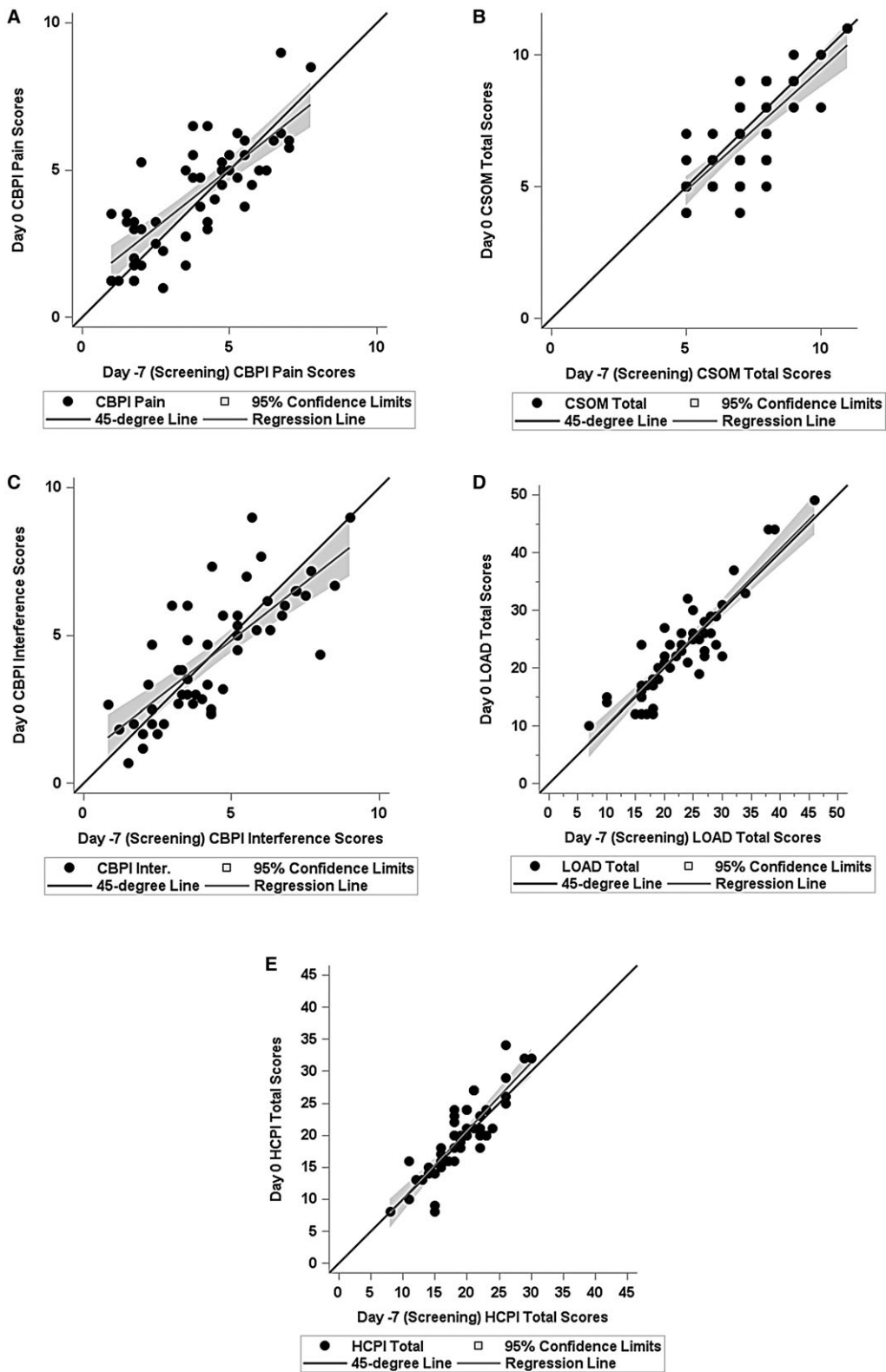
Fig 3. CMI's Baseline Agreement, comparing regression line and 45-degree line. Day −7 CMI scores, on the *x*-axis, are plotted against the Day 0 CMI scores (*y* axis) to determine agreement, with 95% confidence limits represented by gray shading. **(A)** CBPI pain mean baseline scores **(B)**. CSOM total mean baseline scores **(C)**. CBPI interference mean baseline scores **(D)**. LOAD mean baseline scores **(E)**. HCPI mean baseline scores.
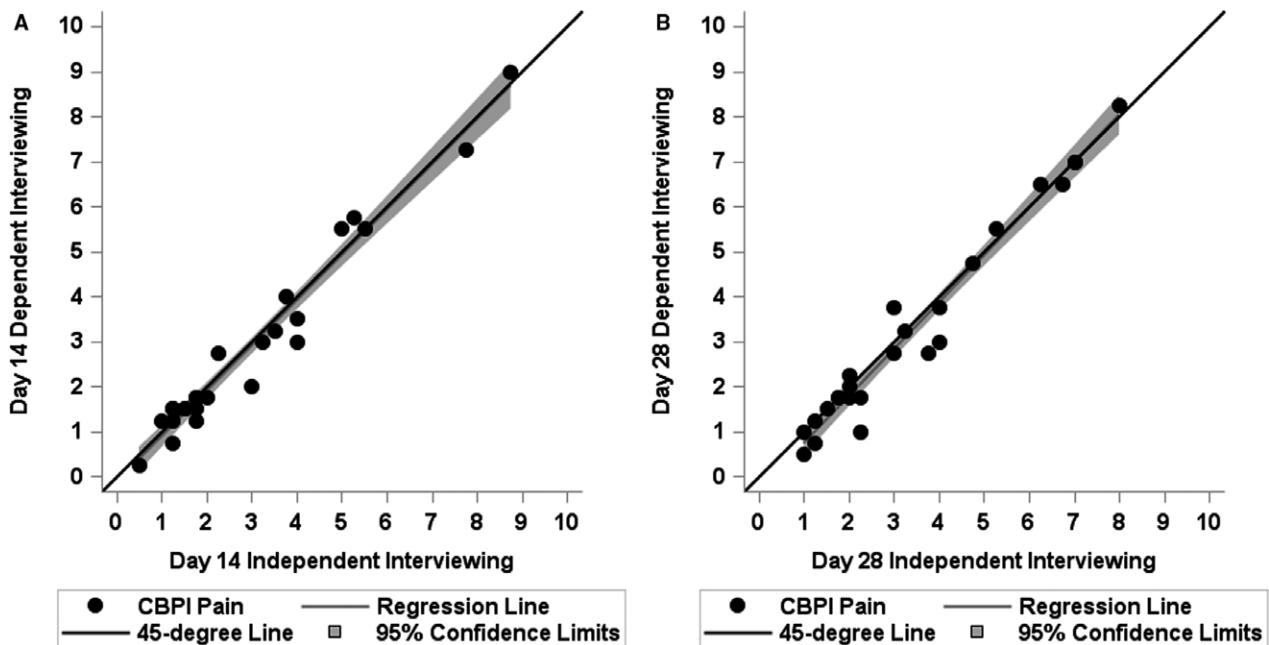
Fig 4. Independent and dependent interviewing CBPI pain scores at Day 14 (A) and at Day 28 (B). The CBPI-independent pain scores, on the x-axis, are plotted against the CBPI-dependent pain scores (y axis), with 95% confidence limits represented by gray shading.
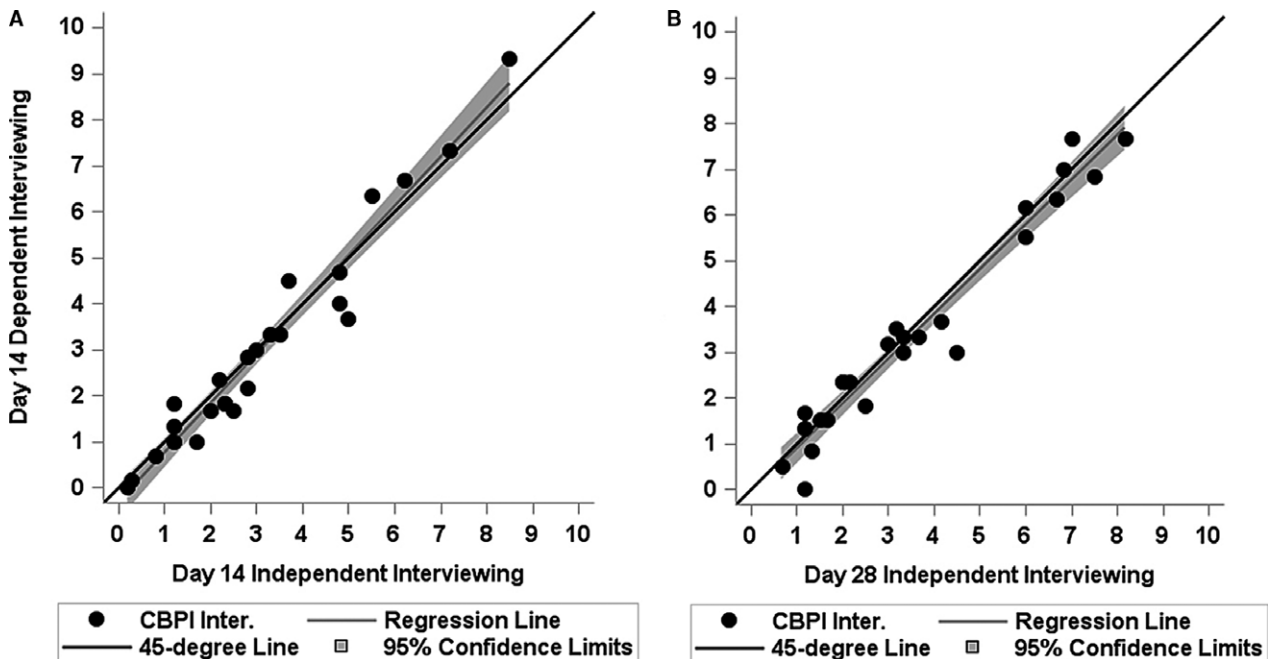

Fig 5. Independent and dependent interviewing CBPI interference scores at Day 14 (A) and at Day 28 (B). The CBPI independent interference scores, on the x-axis, are plotted against the CBPI dependent interference scores (y axis), with 95% confidence limits represented by gray shading.

independent interviewing, and no statistically significant difference between scores collected under the 2 interviewing conditions, leading us to again reject our hypothesis. These data suggest that overall, neither interviewing technique will provide statistically different answers and both appear to be appropriate to use in

the clinical setting. Finally, our data supported our final hypothesis as we found that the use of dependent interviewing provided, on average, lower CMI scores, a higher treatment effect, and a higher effect size. By using dependent interviewing techniques, it could be possible to increase the treatment effect and effect size,
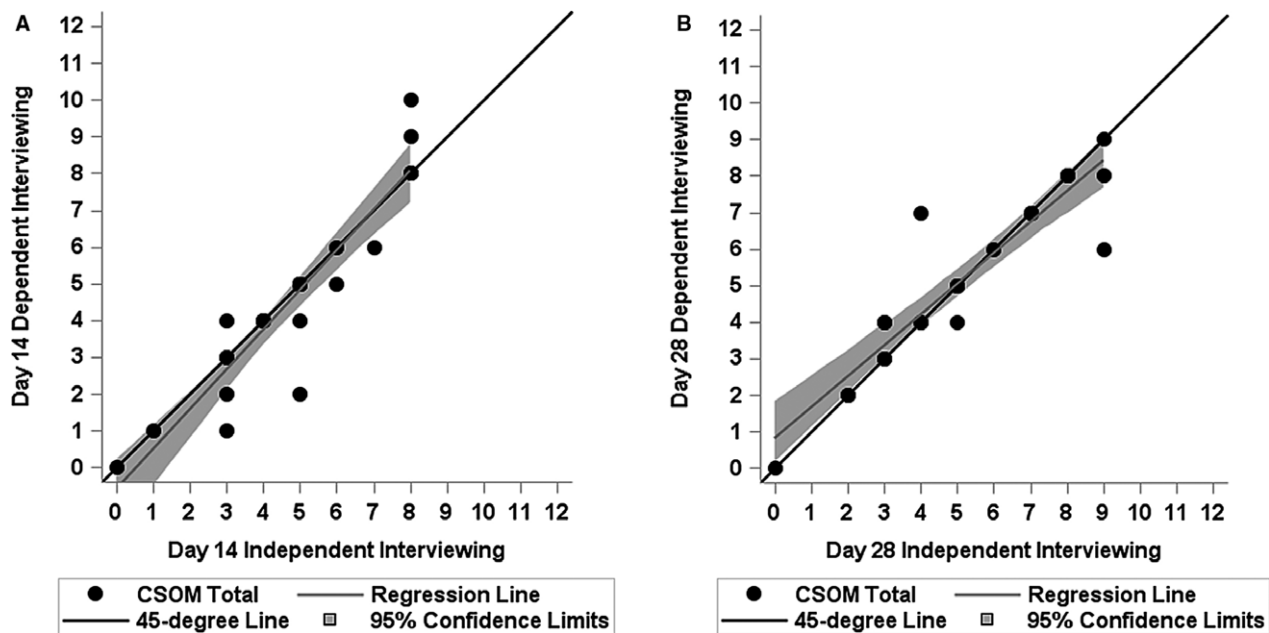
Fig 6. Independent and dependent interviewing CSOM scores at Day 14 **(A)** and at Day 28 **(B)**. The CSOM independent scores, on the x-axis, are plotted against the CSOM dependent scores (y axis), with 95% confidence limits represented by gray shading.

therefore increasing the power of the study through minimal change to study design.

Our first aim was to evaluate baseline variability in CMI scores collected during a 7-day interval, with no treatment having been administered. We had expected that screening CMI scores would decrease between the first (Day −7) and second (Day 0) visit because of the faking bad bias, previously reported by Choi and Aday.[3,12] This appeared to occur for CBPI interference and CSOM Total, with scores indicating significantly less impairment on Day 0 compared to the earlier Day −7, but the opposite was seen for the CBPI pain, LOAD and the HCPI, with the respondent's CMI scores increasing (greater pain and impairment) between Day −7 and Day 0.

A decrease in severity (disability) was expected because of several possible reasons. As discussed previously, the Faking Bad bias may be part of the respondents' discrepancies in the baseline CMI scores; owners may alter their CMI scores (reporting increased pain and impairment) to be enrolled in a study at Day −7. Additionally, the concept of regression to the mean has been discussed[27] where dogs are recruited to a clinical study when the clinical signs are obvious, and natural variation in the disease causes them to become better as the study starts. These 2 phenomena are likely responsible for the generally expected lower scores at a second baseline measurement.

The CBPI interference and CSOM Total CMI results were consistent with our hypothesis and previously reported data. Between Day −7 and Day 0, the CSOM scores significantly decreased in severity. As well as the possible explanations of faking bad and regression to the mean, the CSOM may suffer from another bias. The CSOM uses 3 items tailored to each animal, that

the owners are asked to follow and assess. At Day −7, owners decide on the activity, and rate the difficulty the pet has performing the activity. They then return home, and evaluate the activity. This increased attention to the activity may lead them to realize the pet is not actually as impaired performing the activity as they initially thought, whereas all other CMIs are standardized questionnaires collected using a numeric scale at the time of presentation.

Baseline variability (test-retest reliability) for LOAD, CBPI pain, and HCPI have been previously described,[14,24,25] and all have shown a decrease in CMI scores between the screening (Day −7) and the start of a clinical trial (Day 0), indicating a systemic bias within the baseline screening. However, unlike previous work[24,27] our findings showed only the CBPI interference and CSOM CMI's displayed this trend. All other CMI's (CBPI pain, LOAD, and the HCPI) showed an increase in CMI scores when comparing Day −7 to Day 0 (an increased impairment). This may be caused by a bias called the recall bias or rumination. Recall bias[28] or rumination could cause owners to increase importance of major events or experiences during the clinical trial leading to an alteration of their retest answers. Owners could overinterpret their pet's behavior, causing an increased bias and overrepresentation of their pet's pain and disability.

When evaluating baseline agreement between each CMI, it was noted that the LOAD questionnaire had the best agreement and most consistent results, when comparing Day −7 to Day 0, and may indicate the LOAD is inherently the most stable CMI. The ICC for LOAD found in the current study was identical to the ICC reported by Hercock et al.[24] evaluating dogs with osteoarthritis, confirming its inherent baseline stability.
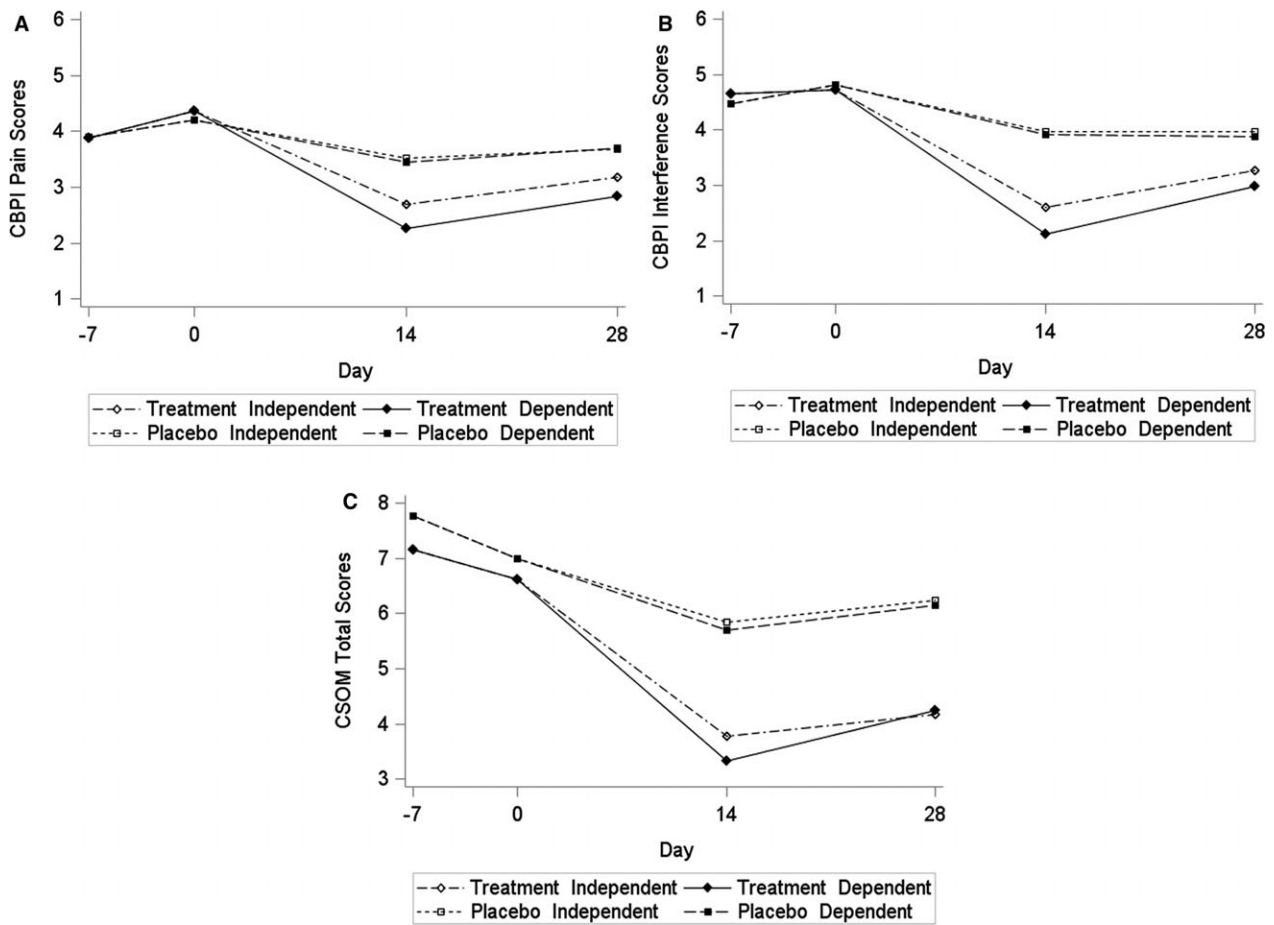
Fig 7. Graphical plots of CMI mean scores over time, showing placebo independent interviewing, placebo dependent interviewing, treatment group independent interviewing, and treatment group dependent interviewing. These plots show demonstrate consistently lower pain scores associated with dependent interviewing when compared to independent interviewing. **(A)** CBPI pain scores **(B)** CBPI interference scores **(C)** CSOM total scores.

Although statistically not different, the variability in baseline scores in all the CMIs suggests that researchers using these instruments should take this into consideration when deciding which baseline time point to use when using "change from baseline" as the primary outcome measure.[14]

Although there was no statistical difference between the 2 survey techniques (independent and dependent interviewing), an interesting pattern was found with a consistent decrease in CMI scores across both the placebo and treatment groups, most pronounced in the treatment group, when using the dependent interviewing technique versus the independent interviewing technique. This phenomenon was also shown within the US Census Bureau's Survey Income and Program Participation (SIPP), where using independent interviewing resulted in increased variability in answers between 2 time points (seam bias) during the longitudinal study, compared to dependent interviewing techniques.[29,30] As a result of this research and the inherent forms of bias presumed to be associated with independent interviewing, the US Census Bureau adapted the dependent interviewing techniques in 2004.[16,18]

We also found that the answers from the dependent interviewing technique showed a greater treatment effect and effect sizes, compared to the independent interviewing technique. These are very interesting findings, and the implications for clinical study design and power are significant. Previous reported methods to increase treatment effects include increasing sample size, reducing measurement error, and raising the alpha level.[31] However, increasing sample size has proven to be costly, and raising the alpha level is not recommended. Using dependent interviewing techniques could increase the treatment effect and effect size, therefore increasing the power of the study through minimal change to study design and no extra cost. However, our results need to be replicated before this approach can be strongly recommended.

We are not aware of any work in veterinary medicine that compares the 2 interviewing approaches. From our data, we cannot conclude whether the independent or dependent interviewing technique provided more accurate results, but given that the difference between the 2 sets of data were more pronounced for the treatment group, it is tempting to think that dependent interviewing

may give a more accurate representation of pain and mobility. Further studies should evaluate the 2 techniques against objective measures of mobility and lameness.

Summary

We identified some baseline variability in CSOM, CBPI, LOAD, and HCPI scores when measured 7 days apart with no intervening treatment, and the direction of shift in baseline scores was inconsistent across all the CMIs, with the best agreement in baseline scores being found for the LOAD CMI, and the least good for the CSOM CMI. Overall, we found good agreement between dependent interviewing and independent interviewing, however, the use of dependent interviewing provided, on average, lower CMI scores, a higher treatment effect, and a higher effect size.

Further research needs to be conducted to determine if these results can be replicated, particularly across a larger sample size and longer timeframe.

Footnotes

Acknowledgments

References

1. Stone DH. Design a questionnaire. BMJ 1993;307:1264–1266.

2. Bowling A. Mode of questionnaire administration can have serious effects on data quality. J Pub Health 2005;27:281–291.

3. Choi BCK, Pak AWP. A catalog of biases in questionnaires. Prev Chronic Dis 2005;2:1–13.

4. Cook C. Mode of administration bias. J Man Manip Ther 2010;18:61–63.

5. Glaser AW, Davies K, Walker D, et al. Influence of proxy respondents and mode of administration on health status assessment following central nervous system tumors in childhood. Qual Life Res 1997;6:43–53.

6. Marx RG, Menezes A, Horovitz L, et al. A comparison of two time intervals for test-retest reliability of health status instruments. J Clin Epidemiol 2003;56:730–735.

7. McColl E, Jacoby A, Thomas L, et al. Design and use of questionnaires: A review of best practice applicable to surveys of health service staff and patients. Health Technol Assess 2001;5:87–88.

8. Schwarz N. Cognitive aspects of survey methodology. Appl Cognit Psychol 2007;21:277–287.

9. Peer E, Gamliel E. Too reliable to be true? Response bias as a potential source of inflation in paper-and-pencil questionnaire reliability. Pract Assess Res Eval 2001;16:1–8.

10. Quelhas A, Santos A, Araujo B, et al. Biases in questionnaire construction: How much do they influence the answers given? [Online]. Available at: http://medicina.med.up.pt/im/trabalhos_10_11/Sites/Turma21/Protocolo%20Final.pdf. 2011. Accessed July, 2015.

11. Last JM, Association IE. A Dictionary of Epidemiology, 5th ed. Oxford: Oxford University Press; 2001.

12. Aday LA, Cornelius LJ. Designing and Conducting Health Surveys: A Comprehensive Guide, 3rd ed. San Francisco, CA: Jossey-Bass; 2006.

13. 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;8:1–10.

14. 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;27:22–30.

15. Lascelles BDX, Hansen B, Roe S, et al. Evaluation of client-specific outcome measures and activity monitoring to measure pain relief in cats with osteoarthritis. J Vet Intern Med 2007;21:410–416.

16. Jäckle A, Lynn P. Dependent interviewing and seam effects in work history data. J Off Stat 2007;23:529–551.

17. Callegaro M. Seam effects in longitudinal surveys. J Off Stat 2008;24:387–409.

18. Moore J, Bates N, Pascale J, et al. Tackling Seam Bias Through Questionnaire Design. Methodology of Longitudinal Surveys. London: John Wiley & Sons, Ltd; 2009:73–92.

19. Jackle A. Dependent Interviewing: A Framework and Application to Current Research. Methodology of Longitudinal Surveys. London: John Wiley & Sons, Ltd; 2009:93–111.

20. Statistics New Zealand. A Longitudinal Survey of Income: Employment and Family Dynamics Feasibility Project Final Report. Wellington: Statistics New Zealand; 2001.

21. Lascelles BDX, Knazovicky D, Case B, et al. A canine-specific anti-nerve growth factor antibody alleviates pain and improves mobility and function in dogs with degenerative joint disease-associated pain. BMC Vet Res 2015;11:101.

22. Baldwin K, Bartges J, Buffington T, et al. AAHA nutritional assessment guidelines for dogs and cats. J Am Anim Hosp Assoc 2010;46:285–296.

23. Brown D, Boston R, Coyne J, et al. Development and psychometric testing of an instrument designed to measure chronic pain in dogs with osteoarthritis. Am J Vet Res 2007;68:631–637.

24. Hercock CA, Pinchbeck G, Giejda A, et al. Validation of a client-based clinical metrology instrument for the evaluation of canine elbow osteoarthritis. J Small Anim Pract 2009;50:266–271.

25. Hielm-Björkman AK, Rita H, Tulamo R-M. Psychometric testing of the Helsinki chronic pain index by completion of a questionnaire in Finnish by owners of dogs with chronic signs of pain caused by osteoarthritis. Am J Vet Res 2009;70:727–734.

26. Valentin S. Cincinnati orthopaedic disability index in canines. Aust J Physiother 2009;55:288.

27. Brown DC, Boston RC, Farrar JT. Use of an activity monitor to detect response to treatment in dogs with osteoarthritis. J Am Vet Med Assoc 2010;237:66–70.

28. Rhodes T, Girman C. Does the mode of questionnaire administration affect the reporting of urinary symptoms? Urology 1995;46:341–345.

29. Martini A. Seam effect, recall bias, and the estimation of labor force transition rates from SIPP. ISER Working Papers 2002:387–392.

30. Moore JC. Seam bias in the 2004 SIPP panel: Much improved, but much bias still remains. Surv Methodol 2008;3: 1–56.

31. Sullivan GM, Feinn R. Using effect size—Or why the P value is not enough. J Grad Med Educ 2012;4:279–282.