

# 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

Anna K. Hielm-Björkman, DVM, PhD; Hannu Rita, PhD; Riitta-Mari Tulamo, DVM, PhD

**Objective**—To determine the validity, reliability, and sensitivity of a published chronic pain index by completion of a questionnaire in Finnish by owners of dogs with chronic signs of pain caused by osteoarthritis.

**Animals**—61 client-owned dogs with osteoarthritis.

**Procedures**—Validity, internal consistency, and repeatability testing of the questionnaire were evaluated by completion of questions in Finnish by owners of 61 dogs; the questionnaire was named the Helsinki chronic pain index (HCPI). Sensitivity testing of the questionnaire was determined from data of 2 smaller groups of dogs treated with carprofen ( $n = 17$ ) or placebo (17). Owners completed the questionnaire 5 times during a 16-week period. Psychometric properties of the HCPI were evaluated.

**Results**—Principal component analysis yielded only a single stable component, indicating that the HCPI was best explained as an 11-item single construct index of chronic pain. Changes in the HCPI correlated well with change in quality of life and with change in the mobility visual analogue scale ( $r = 0.72$  and  $r = 0.67$ , respectively), indicating a high predictive validity. Internal consistency (Cronbach  $\alpha$  value = 0.82) and test-retest reliability (intra-class correlation coefficient = 0.84 for items and 0.92 for the HCPI) were high. Changes in scores (from baseline to treatment and back to no treatment) between the carprofen treatment group and placebo control group were significant, indicating high sensitivity.

**Conclusions and Clinical Relevance**—The Finnish version of the HCPI provided a valid, reliable, and responsive tool for assessment of response to treatment in dogs with osteoarthritis. (*Am J Vet Res* 2009;70:727–734)

Outcome-based veterinary medicine currently lacks reliable validated outcome measures.<sup>1–3</sup> A double-blind, controlled study is seldom used to evaluate outcomes of orthopedic surgery in dogs. On the assumption that chronic pain and lameness are not equivalent in terms of a certain type of behavior but underlie the behavior,<sup>4</sup> chronic pain questionnaires designed to be used by dog owners have been evaluated.<sup>5–8</sup> A simple multifactorial descriptive pain questionnaire in Finnish, leading to an index by summing up scores for 11 questions (ie, items) that were easily applicable to all kinds of dogs, owners, and environments and that all were significantly different in dogs with chronic signs of pain caused by osteoarthritis, compared with healthy nonpainful control dogs, was evaluated by us at the University of Helsinki.<sup>5</sup> We propose to call the resulting index the HCPI.

Received May 8, 2008.

Accepted August 14, 2008.

From the Department of Equine and Small Animal Medicine, Faculty of Veterinary Medicine (Hielm-Björkman, Tulamo), and the Department of Forest Resource Management, Faculty of Agriculture and Forestry (Rita), University of Helsinki, Helsinki, Fin-00014, Finland.

Address correspondence to Dr. Hielm-Björkman.

## ABBREVIATIONS

HCPI	Helsinki chronic pain index
ITT	Intention to treat
PCA	Principal component analysis
QOL	Quality of life
VAS	Visual analogue scale

Before use, it must be determined that an index is valid, reliable, and sensitive.<sup>9</sup> Many tests exist to determine these factors, and no single test can unequivocally prove the worth of an index, but together, test results can strengthen an index.<sup>9</sup> Methods are chosen partly on the basis of how data are gathered and the preferences of the researchers. Part of the validity testing for our index was performed in an earlier study.<sup>5</sup>

Validity is the quality of a scale or index, meaning its ability to measure what it is supposed to measure.<sup>9,10</sup> Face validity<sup>4,10,11</sup> is the extent to which the scale or index after it is constructed is subjectively viewed by knowledgeable individuals (eg, veterinarians) as covering the concept, such that each item in the questionnaire measures chronic pain in some way. Content validity<sup>4,9–11</sup> is related to face validity, being based on logic and expertise. It asks whether the scale or index covers

all of the generally accepted variables of, for example, chronic pain (ie, is it sufficiently comprehensive?). To achieve face and content validity, researchers find the best items that assess chronic pain. This involves a long pretrial process and, in the case of our index, started a year before the final questionnaire was administered.<sup>5</sup> Question topics for the first items were gathered from the clinical experience of the authors, previous research, literature, and informal interviews with owners and colleagues. Questions were tested several times until all ambiguous or poorly worded questions had been deleted or rewritten and again retested. We finally had 25 items that were tested in a clinical setting.<sup>5</sup> From these, 14 items that either were not applicable to all owners (eg, stair climbing), were not easily understood (eg, pacing), or did not reveal a significant difference between healthy and diseased dogs (eg, appetite) were dropped, resulting in an 11-item index.<sup>5</sup> Criterion (also called predictive or concurrent) validity<sup>4,9-11</sup> is used when describing the correlation between a scale and another external measurement of the same phenomenon. In the study reported here, criterion validity was assessed against a QOL question<sup>7,8</sup> and a mobility VAS.<sup>12,13</sup> Construct validity<sup>4,9-11</sup> by extreme groups,<sup>10</sup> in which dogs with pain caused by osteoarthritis were compared with healthy dogs with no pain, was used in our previous study.<sup>5</sup> In the study reported here, construct validity was assessed by use of PCA.<sup>4,14</sup> The primary application of this technique in scale development is to reduce the number of items and to detect a structure in the relationship between items (ie, to determine how many latent constructs underlie a set of items).<sup>4</sup> An important aspect of the construct extraction is that the solution has to be interpretable. Therefore, several solutions are usually explored and the one that makes the best sense is chosen. Usually, the construct extraction is done at a single time point, but if several similar evaluations are at the disposal of researchers, it is possible to check the stability of the construct solution by rechecking it at various time points.<sup>8,10,14,15</sup>

Reliability refers to the extent to which the measure yields the same score each time it is administered, all other things being equal.<sup>9,10,14</sup> When designing a questionnaire, its optimal length is estimated on the basis of how the questionnaire will be used. Kline<sup>15</sup> has recommended 10 items as the minimum for a reliable test. A longer questionnaire is usually more reliable, whereas a shorter questionnaire places less burden on the responder.<sup>4,9</sup> We were aiming for a 10- to 15-item scale with the trade-off of somewhat lower reliability. Several types of reliability<sup>10,15</sup> exist, and the reliability of a scale or an index is often measured by use of different techniques that will give somewhat different values.<sup>9</sup> Internal consistency or equivalence<sup>9</sup> is the first reliability test to perform and estimates how well items that reflect the same construct yield similar results.<sup>14</sup> Internal consistency is assessed through the overall correlation among items in the same construct; the Cronbach  $\alpha$  value<sup>16</sup> is the best-known statistic for this determination. Repeatability (also called stability, test-retest reliability, temporal reliability, or intraobserver reliability) is when a test is given twice to the same cohort and thereby evaluated by the test-retest method.<sup>4,9,10</sup> When the measure is

taken over time intervals, all other things being equal, scores of the owners should remain consistent. This is often tested by use of intraclass correlation.<sup>10,11,17</sup> The correlation is a function of time; 1 month between 2 evaluations will give a value lower than if there would only have been a day or a week between evaluations.

Sensitivity to change or responsiveness<sup>10</sup> of the scale reflects the capability of the instrument to measure changes in degrees of pain over time in response to clinical interventions. As intervention, 1 group receives an analgesic treatment (eg, NSAID for osteoarthritis) and is compared with a group that receives a placebo. The analgesic is presumed to affect the degree of pain more than will the placebo.

The purpose of the study reported here was to validate our previous questionnaire and confirm that psychometric properties of the questionnaire are reliable. We hypothesized that the HCPI we previously developed to quantify owner-assessed chronic pain of dogs with osteoarthritis is valid and reliable. Our method to show validity and reliability was to test the following: construct validity, where component analysis suggests a stable component structure of the HCPI; criterion validity, by comparing a change in the HCPI with 2 other scales that are thought to change because of chronic pain; internal consistency, for which a Cronbach  $\alpha$  value  $> 0.7$  would indicate a correlation among items of the components; repeatability, for which correlations of  $r > 0.7$  would indicate a good test-retest reliability; and responsiveness, for which a significant difference in the HCPI and its items as a result of medication administration (but not without medication administration) would indicate sensitivity of the HCPI to detect change.

## Materials and Methods

**Study design, dog selection, and study procedures**—Data for the present study came from a cohort of 68 dogs participating in a 4-group clinical trial for dogs with chronic signs of pain caused by osteoarthritis.<sup>18,19</sup> The 68 dogs were chosen on the basis of 124 telephone interviews with suitable owners. Inclusion criteria into the clinical trial were that dogs had clinical signs of osteoarthritis and a radiographic diagnosis of moderate or severe osteoarthritis in either a hip joint or an elbow joint. The owner had to describe at least 2 of the following clinical signs as being frequent: difficulty lying down or getting up from a lying position, difficulty jumping or refusing to jump, difficulty walking up or down stairs, or definite lameness. The number of dogs needed for each group was calculated in the clinical trial but not for psychometric testing of the questionnaire.

At 4 weeks before the trial started ( $W_{-4}$ ), some owners were giving their dogs analgesics, but at the beginning of the pain-treatment assessment study ( $W_0$ ), owners were asked to stop administration of all medications to treat pain and osteoarthritis (eg, not to give the dogs NSAIDs, corticosteroids, or pentosan polysulphate sodium<sup>a</sup>). However, 7 of the 68 owners believed that their dogs had too much pain without medication and these owners gave their dogs NSAIDs. The clinical study<sup>18,19</sup> was designed as a randomized double-blind

controlled clinical trial on the basis of published guidelines.<sup>20–28</sup> A person who was blinded to treatments made the first appointments, and at this first visit ( $W_0$ ), dogs were assigned into groups in order of arrival by use of a computer-generated randomized list. Only the location of the disease (hip joint or elbow joint) was stratified for randomization. Radiographs were made of affected hip and elbow joints. Follow-up questionnaires were completed by owners at 4, 8, and 12 weeks ( $W_4$ ,  $W_8$ , and  $W_{12}$ , respectively) for reassessment. Dogs were given study products orally for 8 weeks, from  $W_0$  to  $W_8$ . At  $W_{12}$ , dogs had been without medication for 4 weeks and were evaluated to determine long-term effects of the different treatments.

Products tested in the clinical trial were a green-lipped mussel nutraceutical ( $n = 17$  dogs)<sup>18</sup> and a homeopathic combination preparation (17).<sup>19</sup> Two control groups were included: a positive control group that received carprofen<sup>b</sup> at a dose of 2 mg/kg twice daily ( $n = 17$ ) and a negative control group that received only placebo (17). All 3 test products were supplied by the manufacturers in 2 varieties, which included the actual product to be tested and a placebo that looked identical to the test product; all products were in plain containers with no brand or marking. Products arrived coded and were organized by a research assistant who was not involved in the study thereafter. To perform a blinded study, the 4 groups were medicated as follows: all dogs went home at  $W_0$  with 3 products to be given to them daily for 8 weeks; different combinations of the test products (1 test product/treatment group) and placebos (2 placebos/treatment group and 3 placebos/placebo control group) were provided, depending on the group to which dogs had been allocated. For ethical reasons, all owners were also given a package of 50-mg carprofen<sup>b</sup> tablets in normal packaging at the start of the trial. This rescue analgesic could be used as additional pain relief (dose of 1 tablet for a dog with a body weight of 20 to 30 kg, 2 tablets for a dog with a body weight of 31 to 40 kg, and 3 tablets for a dog with a body weight of 41 to 60 kg) if owners felt their dog was in pain; its use was recorded. Results of the clinical trial are reported elsewhere.<sup>18,19</sup>

Inclusion criteria for the psychometric internal consistency, repeatability, and validity testing of the questionnaire were that owners had answered 2 baseline questionnaires ( $W_{-4}$  and  $W_0$ ), and that their owner-reported medication administration had changed at most by 1 unit on a 5-point medication administration scale (1 = no medication administration during the last 4 weeks, 2 = medication administration 1 to 2 times during the last 4 weeks, 3 = medication administration about once a week, 4 = medication administration about 3 to 5 times in a week, and 5 = daily to almost daily medication administration) between the 2 baselines. Sixty-one dogs fit these criteria and were included in the study.

Inclusion criteria for the sensitivity testing of the questionnaire were different. Results of a sensitivity study determine whether an index is able to detect a difference between 2 groups of dogs that are treated differently. In dogs with osteoarthritis, it is presumed that dogs given placebo will have less response to pain

treatment than dogs given an established analgesic (ie, carprofen).<sup>29</sup> Because 2 of the 4 initial treatment groups were given products considered complementary therapies<sup>18,19</sup> with unknown effects ( $n = 34$  dogs), data on these dogs were not suitable for this sensitivity testing and only data on dogs from the positive control group (ie, dogs that received carprofen; 17) and the negative control group (ie, dogs that only received placebo; 17) were used in this portion of the study and are referred to as the carprofen treatment group and placebo control group, respectively.

All evaluators (veterinarians and owners) and technical assistants were blinded to treatments at all times. Only the statistical analyses in the sensitivity study required that the evaluators be somewhat unblinded, to the extent that data from the 2 groups of dogs that received complementary therapies were excluded.

**Pain assessment questionnaire**—The questionnaire was in Finnish. As previously described, the chronic pain index total score was constructed as the sum of answers to 11 questions.<sup>5</sup> Each answer could be chosen from a 5-point descriptive scale. Answers were later tied to a value (0 to 4) and, when summed, gave a minimum total index score of 0 and a maximum of 44. Values and how to compute the total score were not available to owners while answering the questionnaire.

On the HCPI, owners were asked to check only 1 answer that best described their dog to each of the following 11 statements (as translated from Finnish to English): item 1, rate your dog's mood (very alert [0], alert [1], neither alert nor indifferent [2], indifferent [3], or very indifferent [4]); item 2, rate your dog's willingness to participate in play (very willingly [0], willingly [1], reluctantly [2], very reluctantly [3], or does not at all [4]); item 3, rate your dog's vocalization in the form of audible complaining, such as whining or crying out (never [0], hardly ever [1], sometimes [2], often [3], or very often [4]); items 4 to 7, rate your dog's willingness to walk, trot, gallop, and jump (eg, into car or onto sofa), respectively (very willingly [0], willingly [1], reluctantly [2], very reluctantly [3], or does not walk, trot, gallop, or jump, respectively, at all [4]); items 8 and 9, rate your dog's ease in lying down and in rising from a lying position, respectively (with great ease [0], easily [1], neither easily nor with difficulty [2], with difficulty [3], or with great difficulty [4]); items 10 and 11, rate your dog's difficulty in movement after a long rest and after major activity or heavy exercise, respectively (never [0], hardly ever [1], sometimes [2], often [3], or very often or always [4]).

This questionnaire was answered 5 times at 4-week intervals: twice for baseline ( $W_{-4}$  and  $W_0$ ), twice during treatment ( $W_4$  and  $W_8$ ), and once at follow-up ( $W_{12}$ ). Both baseline evaluations took place during the dry, cold winter season with the temperatures being similar as a means to minimize the influence of weather and temperature on the evaluated osteoarthritic pain.

At  $W_4$ , owners also answered a relative question on their dog's overall QOL, in which a comparison was made between the QOL at that time to that at  $W_0$ : 5 standard responses were provided (1 = much better, 2 = better, 3 = the same, 4 = worse, and 5 = much worse). At  $W_0$  and  $W_4$ , owners also evaluated lameness by use of a

VAS. The VAS consisted of a 100-mm-long line with the left end point designated as sound (ie, no difficulties in locomotion) and the right end point designated as most lame (ie, most severe difficulties in locomotion possible); owners were asked to mark the perceived mobility of their dog by drawing a cross on the line.

Owners completed the questionnaire on the behavior of their dog as proxies, at home, without any prior advice or instructions, and returned the questionnaire to the clinic. All observers or evaluators were owners or at least living in the same household as the dog. Owners were all naive to acting as a proxy and had never done this kind of evaluation. No data were collected on owner characteristics. Owners of the dogs were required to sign informed consent forms. The study protocol was approved by the Ethics Committee of the University of Helsinki.

**Statistical analysis**—To minimize the postrandomization selection bias, data from all eligible dogs were included by use of ITT analysis.<sup>30</sup> Construct validity was studied at  $W_0$  by use of PCA with and without varimax rotation. A Keiser-Mayer-Olin measure<sup>31</sup> of sampling adequacy  $> 0.6$  indicated that data were suitable for component analysis. Only constructs with an eigenvalue  $> 1$  were interpreted,<sup>14</sup> and loading values  $> 0.4$  were emphasized.<sup>10</sup> From the scree plot, Cattell<sup>32</sup> suggested that all components located on the horizontal portion be discarded, except if a distinct elbow was observed between the vertical and horizontal portions, in which case these data are also extracted. The component structure that was most interpretable was suggested as the structure of the HCPI. The interitem correlation matrices for the HCPI items are reported. Criterion validity was analyzed by correlating the change in the HCPI between  $W_0$  and  $W_4$  to a change in QOL and to a change in a mobility VAS by use of the Spearman rank test. Internal consistency or the degree of mean correlation at  $W_0$  among the 11 items of the HCPI and among items of the components extracted was evaluated by use of the Cronbach  $\alpha$  value. The test-retest reliability model was used when the mean values of the HCPI and the 11 individual items of the questionnaire were compared at 2 baseline measurements ( $W_{-4}$  and  $W_0$ ) by use of intraclass correlation tests. A high intraclass correlation and similarity of the mean values indicated repeatability.

Sensitivity to change of the HCPI was studied by use of the 2-independent-sample Mann-Whitney test to compare HCPI values and their changes between the carprofen treatment group and placebo control group separately at and between various time points.<sup>10</sup> In the ITT analyses, data from all dogs that came in for an evaluation were used, regardless of whether dogs deviated by use of rescue medication administration or not. Baseline bias between the carprofen treatment group and placebo control group was assessed by use of a  $\chi^2$  test and cross tabulation. Lower index scores indicated less pain, and higher scores indicated more pain. Age, duration of signs, and HCPI baseline values were controlled for in analyzing treatment efficacy. Lowered values of the total index score in the carprofen treatment group during medication administration indicated the sensitivity of the HCPI in response to change. Analy-

sis of changes in scores comparing the carprofen treatment group and placebo control group was used. All tests were 2 tailed, and significance was set at a value of  $P < 0.05$ . Controlling for variables with baseline bias was done by use of a software program<sup>c</sup>; all other tests and calculations were performed by use of a different software program.<sup>d</sup>

## Results

Inclusion criteria for internal consistency, repeatability, and validity testing of the questionnaire resulted in retention of data from 61 dogs. For 9.8% (6/61) of dogs, a 1-step change in medication administration occurred between the 2 baseline evaluations. Baseline data for the 61-dog cohort were as follows: 55.7% (34/61) males, 44.3% (27/61) females, median age of 6 years (range, 1 to 11 years), median body weight of 34 kg (range, 20 to 60 kg), and median time with clinical signs of pain of  $> 2$  years (range, 2 to 6 months to  $> 2$  years). Twenty-four breeds were represented as follows: German Shepherd Dogs ( $n = 15$  dogs), Rottweilers (5), Boxers (4), Golden Retrievers (4), Newfoundlands (4), Samoyeds (4), and all other breeds (with 1 to 3 dogs/breed). There was only 1 mixed-breed dog.

**Construct validity**—The Keiser-Mayer-Olin measure<sup>31</sup> of sampling adequacy was equal to 0.78, indicating that the data were suitable for component analysis. Principal component analysis loading and interitem correlation values were determined (Table 1). The unrotated PCA matrix at  $W_{-4}$  and  $W_0$  resulted in extraction of 3 components with an eigenvalue  $> 1$ . However, according to the interpretation of Cattell<sup>32</sup> for scree plots, only the first (component 1) should be extracted (Figure 1) from  $W_0$ . This component contained 10 strong items with similar unrotated loading at both baseline evaluations ( $W_{-4}$  and  $W_0$ ; 0.44 to 0.70 and 0.44 to 0.76, respectively) and 1 weaker item, vocalization, with loading of only 0.20 and 0.27, respectively (Figure 2). In this single construct solution, all 11 items could be interpreted as variables indicative of chronic pain and a high correlation ( $r = 0.99$ ) was found at  $W_0$  between component 1 and the HCPI value. The mutual correlation ( $r = 0.91$ ) of component 1 at  $W_{-4}$  and  $W_0$  was also high.

When construct validity was further evaluated by a varimax rotation at  $W_{-4}$  and  $W_0$ , 3 components with an eigenvalue  $> 1$  were extracted. However, at  $W_{-4}$ , the components were not interpretable in a meaningful manner, whereas the rotation at  $W_0$  resulted in extraction of 3 components that could be clinically interpreted. These 3 components at  $W_0$  explained 59.1% of the total variation among the 11 index items. In the first component, 8 items (items 4 to 11) were identified with high component loading (ie, 0.50 to 0.80); because items were all related to mobility, this component was referred to as the mobility component. The second component had 2 items related to mood (items 1 and 2) with component loading of 0.78 to 0.92. The third component had 1 item, vocalization (item 3), with a loading of 0.88. Because items loaded differently at  $W_{-4}$  and  $W_0$ , the varimax rotation at  $W_{12}$  was also evaluated to get a clearer understanding of how to best pres-

Table 1—The HCPI values (mean  $\pm$  SD) for items totaled and individual items with the corresponding PCA loading and interitem correlation values and Cronbach  $\alpha$  values (n = 61).

HCPI items*	HCPI values						
	$W_{-4}$	$W_0$	Difference of means ( $W_0 - W_{-4}$ )	PCA loading (r)†	Communalities (h²)‡	Item-total correlation (r)§	Cronbach $\alpha$ value for HCPI
Items totaled	16.05 $\pm$ 5.58	15.77 $\pm$ 5.52	-0.278	NA	NA	NA	0.82
Mood	1.02 $\pm$ 0.76	0.95 $\pm$ 0.72	-0.066	0.66	0.86	0.53	0.80
Play and games	0.87 $\pm$ 0.87	0.92 $\pm$ 0.76	0.049	0.63	0.66	0.49	0.81
Vocalization	0.77 $\pm$ 0.94	0.75 $\pm$ 0.89	-0.016	0.27	0.80	0.21	0.83
Walking	1.13 $\pm$ 0.83	1.10 $\pm$ 0.79	-0.033	0.76	0.58	0.66	0.79
Trotting	1.54 $\pm$ 1.00	1.43 $\pm$ 0.94	-0.115	0.44	0.30	0.33	0.82
Galloping	1.33 $\pm$ 0.94	1.38 $\pm$ 0.97	0.049	0.66	0.43	0.56	0.80
Jumping	1.90 $\pm$ 1.09	1.80 $\pm$ 1.05	-0.098	0.54	0.41	0.43	0.81
Lying down	1.66 $\pm$ 0.73	1.66 $\pm$ 0.73	0.000	0.72	0.55	0.62	0.80
Getting up	2.26 $\pm$ 0.84	2.18 $\pm$ 0.85	-0.082	0.74	0.68	0.64	0.79
Movement after rest	2.66 $\pm$ 1.02	2.59 $\pm$ 0.76	0.098	0.56	0.63	0.46	0.81
Movement after major exercise	0.92 $\pm$ 0.76	1.02 $\pm$ 0.70	-0.066	0.69	0.59	0.57	0.80

\*Value for items totaled ranged from 0 to 44; values for individual items ranged from 0 to 4. †PCA loading values are correlations between items and the component; values > 0.4 indicate that they are highly correlated. ‡Communalities are proportions of variance for each item that can be explained by the component; values > 0.4 indicate that the item is related to the other items. §Item-total correlations are correlations between individual items and the HCPI (when that item is omitted); items with values > 0.2 may be retained. || Cronbach  $\alpha$  value measures the extent to which the item responses are correlated to each other;  $\alpha$  values > 0.7 indicate that items can be considered parts of this scale.  
NA = Not applicable.

ent the HCPI. However, in this instance, items loaded differently; component 1 was involved with signs of chronic pain associated with active movement (items 1, 2, 4 to 7, and 11), component 2 was linked to signs of chronic pain associated with lying down (items 8 to 10), and component 3 was associated with vocalization (item 3).

**Criterion validity**—Changes in HCPI significantly correlated with changes in QOL ( $r = 0.72$ ;  $P < 0.001$ ) and lameness ( $r = 0.67$ ;  $P < 0.001$ ).

**Internal consistency**—The Cronbach  $\alpha$  value for the 11 questions at time  $W_0$  was 0.82 (Table 1), with an interitem correlation mean of 0.31, indicating internal consistency with an acceptable degree of mean correlation among questions. If single items were deleted, the Cronbach  $\alpha$  value ranged from 0.79 to 0.83. If the HCPI at  $W_0$  was evaluated on the basis of 3 components, the Cronbach  $\alpha$  value for the first mobility component was 0.81 and the Cronbach  $\alpha$  value for the second mood component was 0.80, whereas the Cronbach  $\alpha$  value for the third vocalization component was not calculated because the component included only 1 item.

**Repeatability**—The HCPI and the 11 items had intraclass correlations of 0.92 and 0.84, respectively, when tested at the 2 baseline evaluations that were 4 weeks apart (ie, at  $W_{-4}$  and  $W_0$ ), indicating a high test-retest reliability. At the 2 baseline evaluations, mean

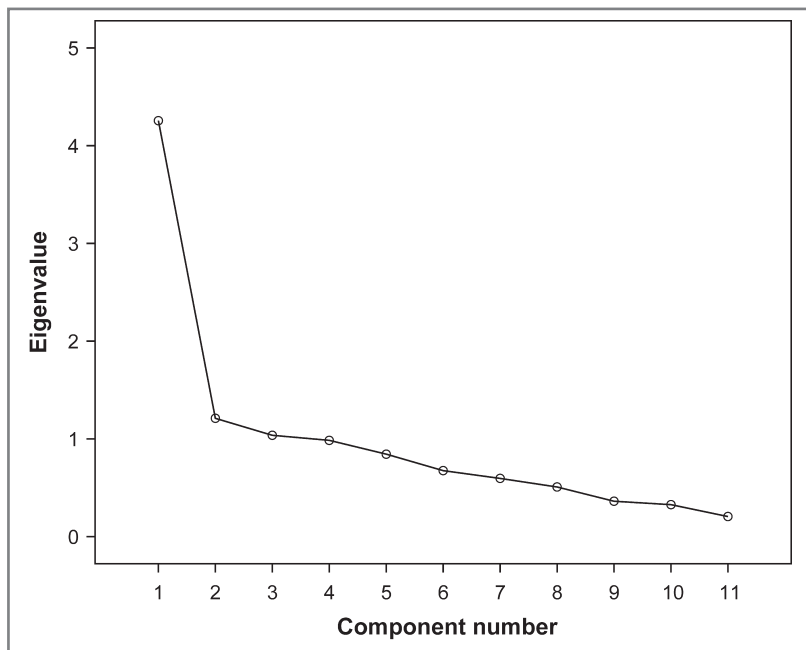


Figure 1—Scree plot of the eigenvalues of the HCPI at  $W_0$ . The shape of the graph indicates that the HCPI is best explained as a single component index, including all of the 11 items (n = 61). This component is referred to as a general chronic pain component.

HCPI total scores were similar; the observed change in mean HCPI total score from  $W_{-4}$  to  $W_0$  was only 0.278, which is < 2% of the mean index value at  $W_{-4}$  (16.05; Table 1).

**Sensitivity to change**—No significant baseline differences were found between the 2 treatment groups in terms of HCPI total score, breed distribution, osteoar-

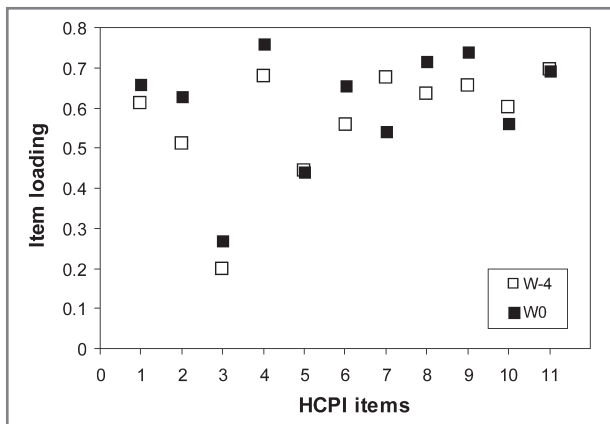


Figure 2—Loading values of the individual 11 items of the first principal component at the 2 baseline evaluations 4 weeks before trial start ( $W_{-4}$ ) and at trial start ( $W_0$ ). The linear picture indicates that the HCPI explains different aspects of the same phenomenon (ie, chronic pain) quite well (vocalization less than the others). The closeness of the values at the 2 time points indicate that the responders responded similarly, although there was a 4-week break between the 2 evaluations ( $n = 61$ ). Item 1 = Rate your dog's mood. Item 2 = Rate your dog's willingness to participate in play. Item 3 = Rate your dog's vocalization in the form of audible complaining. Item 4 = Rate your dog's willingness to walk. Item 5 = Rate your dog's willingness to trot. Item 6 = Rate your dog's willingness to gallop. Item 7 = Rate your dog's willingness to jump. Item 8 = Rate your dog's ease in lying down. Item 9 = Rate your dog's ease in getting up from a lying position. Item 10 = Rate your dog's difficulty in movement after a long rest. Item 11 = Rate your dog's difficulty in movement after major activity or heavy exercise.

thrititis location (ie, forelimb or hind limb), sex, or body weight. However, significant baseline differences were found between the 2 treatment groups for 2 variables, age and duration of clinical signs of osteoarthritis. Younger dogs with a longer history of signs of pain were found in the carprofen treatment group (mean age, 5.1 years with > 2 years of clinical signs of osteoarthritis), compared with the placebo control group (mean age, 7.1 years with 1 to 2 years of clinical signs of osteoarthritis).

Median HCPI total scores in the placebo control group and carprofen treatment group were similar at  $W_{-4}$ ,  $W_0$ , and  $W_{12}$ . During the time of medication administration (ie, carprofen in the carprofen treatment group and placebo in the placebo control group), the median HCPI in the placebo control group exceeded that of the carprofen treatment group (indicating that the placebo control group was rated as having more pain than the carprofen treatment group) at  $W_4$  and  $W_8$ , and differences between the 2 groups at these time points were significant ( $P = 0.004$  and  $P < 0.001$ , respectively). When analyzing the changes in scores (from baseline to treatment and back to no treatment) between the carprofen treatment group and placebo control group, differences in change were also significant ( $P = 0.004$  and  $P = 0.002$ , respectively), again indicating that the HCPI is sensitive. When controlling for age, duration of clinical signs, or baseline HCPI, all results were similar and significant findings remained.

## Discussion

Our results indicate that the HCPI in Finnish is usable, valid, and reliable. In this study, the HCPI

was tested for constructs and different solutions were examined. Our conclusion was that the HCPI is best interpreted as a single construct questionnaire, the construct being chronic pain, including all 11 items previously introduced.<sup>5</sup> Several factors led to this conclusion. First, the first component retained accounted for 39% of the variance but the next 10 components accounted for 2% to 11% each, in a linear pattern (Figure 1). This indicated it was better to retain only 1 component. Second, the rule of Cattell<sup>32</sup> indicates that only 1 component be retained on the basis of the shape of our scree plot. Third, as we had data from various time points, of which at least  $W_{-4}$ ,  $W_0$ , and  $W_{12}$  should be quite similar (because no medication was administered at these times), we could see that the items did not load similarly at these 3 time points. At  $W_{-4}$ , none of the 3 components with an eigenvalue > 1 appeared clinically relevant; however, at  $W_0$ , 3 components with an eigenvalue > 1 were clinically interpretable. At  $W_{12}$ , 3 components with an eigenvalue > 1 were also clinically interpretable, but these components were different from those at  $W_0$ . Each item should load on (ie, be correlated with) the scale it belongs to and not on any other scale.<sup>10</sup> Also, it is important that the components make sense. That the loading values changed so much at the various time points possibly indicates that the owners undergo a learning curve and look at different things at different times. It is also possible that it is the undulating nature of osteoarthritis or differences in weather that cause these variations. Because loading values on the first component were high and similar for all items at all of the 3 time points (ie,  $W_{-4}$ ,  $W_0$ , and  $W_{12}$ ), we believe that these items should be considered as explaining a single latent construct, chronic pain. However, vocalization (ie, audible complaining) reacted with less loading than the other items at each time but was still consistent, as it reacted similarly at  $W_{-4}$  and  $W_0$  (Figure 2). This could indicate that this feature is actually not indicative of pain but still consistent. Because this item does not seem to strengthen the validity, one could argue for expelling it from the HCPI, but if deleted, it would only have increased the internal consistency from 0.82 to 0.83. Veterinarians do not recognize vocalization by dogs as a typical sign of chronic pain. However, because whimpering and crying out are typical signs of pain in humans, it is usually the easiest for owners to recognize. Because of this, we did not want to exclude it from our index. The wording of this question made it easy for responders with all kind of dogs to answer it, and dogs that did not vocalize in response to pain simply got a lower index value throughout evaluations.

Item choice, language, and wording of questions are variables that can affect the use and usefulness of an index or scale and that also affect its psychometric properties. It is always a challenge to decide which items to keep in an index. For example, we decided to exclude stair climbing at a primary stage<sup>5</sup> because many dog owners left the question unanswered and later reported that they never saw their dogs climb stairs. Brown et al<sup>8</sup> decided to use a similar item (ability to climb stairs) in their pain inventory, seemingly having no problems with their population on this point. Nevertheless, in-

cluding this kind of hard-to-answer item into a scale might put the inventory at risk if the question is either not answered at all or answered at only some time points, resulting in summed scores that are not comparable, giving inaccurate results on change over time. Also, words might have another meaning or connotation tied to them in another culture, rendering an item unusable. As even dog owner house architecture can have an impact on how useful a questionnaire can be for a researcher, we realize that it is hard, if not impossible, to find a chronic pain index in dogs that would be universal, fit all kinds of dogs in all kinds of environments, and be exactly translatable to all languages and still be valid and reliable.

The psychometric tests reported here are only valid for the Finnish version of the HCPI. However, the first step to validate a questionnaire from another language is having it adequately translated. This has now been done for the HCPI, which has been translated from Finnish to English and back-translated by independent official bilingual translators with medical backgrounds. The English version of the HCPI will have to undergo the same psychometric tests that are presented here to determine that it is also valid and reliable and that it can be understood and used repeatedly by dog owners. A broad index such as the HCPI that purports to assess chronic pain could, apart from different languages, later also be tested and validated for different causes of chronic pain, for different dog breeds, and possibly even for different groups of dog owners.

Criterion validity of the HCPI was tested by correlating these results to those of other already established chronic pain scales. In this study, the HCPI was correlated with 2 other pain scales, QOL assessment and a VAS to assess lameness. Brown et al<sup>8</sup> successfully validated their index against an owner assessment of QOL. However, in a recent study,<sup>33</sup> a VAS used to evaluate dogs with acute experimentally induced lameness did not correlate well with ground reaction forces, possibly indicating that the correlation of owner VAS results with HCPI values as a means of validation may not have been advisable.

Three ways of evaluating reliability (internal consistency, repeatability, and sensitivity to change) of the HCPI were used in the present study, and all indicated a high reliability. Also, the correlations between the 2 baseline evaluations were large for the individual items. If the time interval between baseline measurements is too long, the clinical pain status might change, and if it is too short, the owner might remember prior answers.<sup>10</sup> This study was part of an outcome assessment study in which 4-week intervals were used. Our results indicate that the questionnaire can be given twice at 4-week intervals, and the owners will give nearly the same answers (Table 1). But in reality, pain status for some dogs might change even from day to day, depending on weather, extra activity, and other events, making any interval too long. The second time point ( $W_0$ ) was used for many of the statistical analyses in this study because the first time a new observational questionnaire is used in any cohort, the responders might still be unsure of their own opinion about items asked, as they might not yet have spent time really inspecting the

dog. Therefore, when owners answer the questionnaire for the second time, their responses are probably more accurate.

The sensitivity of the HCPI to react to changes was determined to be good. Results of other studies<sup>29,34</sup> have revealed the analgesic effect of carprofen in dogs with chronic signs of pain, and the results of the present study support these conclusions. As expected, a significant difference was found when comparing the HCPI and its items between the carprofen treatment group and placebo control group during the time of the analgesic or placebo treatment (from  $W_0$  to  $W_8$ , at the 2 evaluations  $W_4$  and  $W_8$ ). No such difference between groups was seen before or after the treatment period. This indicates that the analgesic was effective when taken and that the questionnaire and its index were sensitive enough to measure the difference in mood and mobility of dogs as observed by their owners.

When testing an index, biases that are introduced into the data must be considered. Because this study was randomized, double-blind, and placebo controlled, a minimum of bias should have been present. To get usable groups for the sensitivity study, only dogs given carprofen or placebo were included. Choosing only dogs that would have been best suited for this trial (ie, dogs that did not receive rescue analgesia) would have eliminated 2 to 4 dogs/time point, and as this could have been considered postrandomization selection bias, it was avoided by reporting on all dogs by use of ITT analysis.

In conclusion, the Finnish version of the HCPI is an easy-to-use, reliable, and valid single component questionnaire for evaluation of pain in dogs. The strength of this index is that it comprises 11 questions that all belong to the same construct, that of chronic pain in dogs. Because the HCPI even in this small group of dogs was capable of detecting a clear change in pain during analgesic administration, we propose that the HCPI be used as a tool for chronic pain evaluation in clinical research, in which owners evaluate the outcome of treatments of dogs for osteoarthritis.

- a. Cartrophen vet. injectable 100mg, Biopharm Pty Ltd, Bondi Junction, NSW, Australia.
- b. Rimadyl, 50-mg tablets, Pfizer, Espoo, Finland.
- c. StatXact-8, Cytel Software Corp, Cambridge, Mass.
- d. SPSS, version 12.0, SPSS Inc, Chicago, Ill.

## References

1. Schulz KS, Cook JL, Kapatkin A, et al. Commentary. Evidence-based surgery: time for change. *Vet Surg* 2006;35:697–699.
2. Cook JL. Outcomes-based patient care in veterinary surgery: what is an outcome measure? *Vet Surg* 2007;36:187–188.
3. Kapatkin AS. Outcome-based medicine and its application in clinical surgical practice. *Vet Surg* 2007;36:515–518.
4. DeVellis RF. *Scale development—theory and applications*. Vol 26. 2nd ed. Thousand Oaks, Calif: Sage Publications Inc, 2003.
5. Hielm-Björkman AK, Kuusela E, Liman A, et al. Evaluation of methods for assessment of pain associated with chronic osteoarthritis in dogs. *J Am Vet Med Assoc* 2003;222:1552–1558.
6. Hudson JT, Slater MR, Taylor L, et al. 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.
7. Wiseman-Orr ML, Scott EM, Reid J, et al. Validation of a structured questionnaire as an instrument to measure chronic pain in

- dogs on the basis of effects on health-related quality of life. *Am J Vet Res* 2006;67:1826–1836.
8. Brown DC, Boston RC, Coyne JC, 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.
  9. Carmines EG, Zeller RA. *Reliability and validity assessment*. Thousand Oaks, Calif: Sage Publications Inc, 1979;9–70.
  10. Streiner DL, Norman GR. *Health measurement scales: a practical guide to their development and use*. 2nd ed. New York: Oxford University Press, 1995;4–161.
  11. Nunnally JC. *Psychometric theory*. 2nd ed. New York: McGraw-Hill Book Co, 1978;86–113, 225–325.
  12. Welsh EM, Gettinby G, Nolan AM. Comparison of a visual analogue scale and a numerical rating scale for assessment of lameness, using sheep as a model. *Am J Vet Res* 1993;54:976–983.
  13. Quinn MM, Keuler NS, Lu Y, et al. Evaluation of agreement between numerical rating scales, visual analogue scoring scales, and force plate gait analysis in dogs. *Vet Surg* 2007;36:360–367.
  14. Tabachnick BG, Fidell LS. *Principal components and factor analysis*. In: *Using multivariate statistics*. 5th ed. Boston: Allyn and Bacon, 2007;607–675.
  15. Kline P. Psychometric theory and method. In: *The handbook of psychological testing*. London: Routledge, 1993;5–170.
  16. Cronbach LJ. Coefficient  $\alpha$  and the internal structure of tests. *Psychometrika* 1951;16:297–334.
  17. Cohen J. *Statistical power analysis for the behavioral sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates Inc, 1988.
  18. Hielm-Björkman A, Tulamo R-M, Salonen H, et al. Evaluating a complementary therapy for moderate to severe canine osteoarthritis. Part I: green-lipped mussel (*Perna canaliculus*). *Evid Based Complement Alternat Med* 2007;doi: 10.1093/ecam/nem136.
  19. Hielm-Björkman A, Tulamo R-M, Salonen H, et al. Evaluating a complementary therapy for moderate to severe canine osteoarthritis. Part II: a homeopathic combination preparation (Zeel). *Evid Based Complement Alternat Med* 2007;doi: 10.1093/ecam/nem143.
  20. Budberg SC. The randomized clinical trial. *Vet Surg* 1991;20:326–328.
  21. Begg C, Cho M, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 1996;276:637–639.
  22. Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996;17:1–12.
  23. Farrar JT, Portenoy RK, Berlin JA, et al. Defining the clinically important difference in pain outcome measures. *Pain* 2000;88:287–294.
  24. Akhtar-Danesh N. A review of statistical methods for analyzing pain measurements. *Eur J Pain* 2001;5:457–463.
  25. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *Lancet* 2001;357:1191–1194.
  26. MacPherson H, White A, Cummings M, et al. Standards for reporting interventions in controlled trials of acupuncture—the STRICTA recommendations. *Acupunct Med* 2002;20:22–25.
  27. Asai T. Confidence in statistical analysis. *Br J Anaesth* 2002;89:807–810.
  28. Turk DC, Dworkin RH. What should be the core outcomes in chronic pain clinical trials? *Arthritis Res Ther* 2004;6:151–154.
  29. Holsinger RH, Parker RB, Beale BS, et al. The therapeutic efficacy of carprofen (Rimadyl-V) in 209 clinical cases of canine degenerative joint disease. *Vet Comp Orthop Traumatol* 1992;5:140–144.
  30. Brown DC. Sources and handling of losses to follow-up in parallel-group randomized clinical trials in dogs and cats: 63 trials (2000–2005). *Am J Vet Res* 2007;68:694–698.
  31. Kaiser HF. The application of electronic computers to factor analysis. *Educ Psychol Meas* 1960;20:141–151.
  32. Cattell RB. A guide to statistical techniques. In: *The scientific use of factor analysis in behavioral and life sciences*. New York: Plenum Press, 1978;17–32.
  33. Waxman AS, Robinson DA, Evans RB, et al. Relationship between objective and subjective assessment of limb function in normal dogs with an experimentally induced lameness. *Vet Surg* 2008;37:241–246.
  34. Vasseur PB, Johnson AL, Budberg SC, et al. Randomized, controlled trial of the efficacy of carprofen, a nonsteroidal anti-inflammatory drug, in the treatment of osteoarthritis in dogs. *J Am Vet Med Assoc* 1995;206:807–811.