Application of a scaling model to establish and validate an interval level pain scale for assessment of acute pain in dogs

Carolyn M. Morton, BVMS, MVM; Jacky Reid, BVMS, PhD; E. Marion Scott, PhD; Lynne L. Holton, PhD; Andrea M. Nolan, MVB, PhD

Objective—To establish interval level measurement in a prototype composite measure pain scale (CMPS) for assessment of acute pain in dogs and to investigate the scale's validity.

Animals—20 clinically normal dogs, 20 dogs with medical conditions, and 117 dogs undergoing surgery.

Procedure—First, a scaling model was applied to the CMPS descriptors to establish weights for each and create a continuous scale. Subsequently, 5 observers independently used the scale to score signs of pain in 4 groups of dogs (control dogs, dogs with medical conditions, and 40 dogs undergoing soft tissue or orthopedic surgery). Scores from each group and from groups of conditions perceived to cause no, mild, moderate, and severe pain were compared. In addition, the scale was applied to 77 dogs undergoing orthopedic or soft tissue surgery and scores were compared with simultaneously derived numeric rating scale (NRS) scores; comparisons were made between surgical groups and with time after surgery.

Results—Calculated scale descriptor weights ranged from –2.0 to 2.0 and were transformed to create a continuous scale from 0 to 10. Median CMPS scores differed significantly among the 4 study groups and among pain severity groups and were typically greater with increasing perceived pain severity. Agreement was determined between CMPS and NRS scores, and there was a significant and expected time effect and difference between the CMPS scores of dogs undergoing orthopedic and soft tissue surgery.

Conclusions and Clinical Relevance—Results indicate that this interval level measurement scale is a valid measure of acute pain in dogs. (*Am J Vet Res* 2005;66:2154–2166)

Although previously a neglected area, it is now well recognized in veterinary medicine that the provision of optimal patient care includes the management of pain. The ability to recognize and quantify the signs of pain is paramount to the development of effective analgesic strategies. Pain in humans has been defined as an "unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage,"[1] from which has been extrapolated the following definition for animals: "Pain in animals is an aversive sensory and emotional experience (a perception), which elicits protective motor actions, results in learned avoidance, and may modify species-specific traits of behavior, including social behavior. Pain depends on the activation of a discrete set of receptors and neural pathways and is usually or potentially noxious, for example damaging to tissue."[2]

In veterinary medicine to date, methods of scoring pain in domestic animals have been restricted mainly to the use of 3 unidimensional, subjective rating scales that were originally developed for use in humans: the **simple descriptive scale (SDS)**, the **numeric rating scale (NRS)**, and the **visual analog scale (VAS)**. However, these have been shown to be unreliable in veterinary settings,[3] and because they are purely ordinal in nature, they do not provide the necessary discriminatory properties (interval level measurement) required for assessment of modern analgesia.

The measurement of pain, even in humans who can self-report, is complicated by the individual, multidimensional nature of the pain experience. The McGill pain questionnaire[4] was developed to provide a quantitative measure of clinical pain that could be treated statistically and would also take account of the sensory and affective qualities of pain in addition to intensity. Pain measurement in groups of humans who are incapable of self-reporting (eg, neonates, infants, people with severe learning disabilities, or demented and verbally handicapped patients) is even more challenging. The inability of the patient, human or other animal, to accurately convey feelings of pain and suffering increases the duty of care of those who undertake this responsibility. It is in this area that medical and veterinary clinicians share the challenge of producing valid, reliable, and reproducible pain assessment tools for patients incapable of self-reporting.

Most work in humans who are incapable of self-reporting has been undertaken in neonates and babies, for which the scales rely heavily on the composite measurement of behavior and facial expressions. The properties required of such scales have been well documented by psychometricians; however, it has been suggested that these requirements are seldom observed during the construction of scales for use in a medical setting.[5] In veterinary medicine, attempts have been

Received January 26, 2005.
Accepted March 25, 2005.
From the Institute of Comparative Medicine, University of Glasgow, Glasgow G61 1QH, UK (Morton, Reid, Nolan); the Department of Statistics (Scott), University of Glasgow, Glasgow G12 8QQ, UK; and the Royal Bank of Scotland, Gogarburn, Edinburgh EH12 1HQ, UK (Holton).
Funded by Pfizer Animal Health and British Small Animal Veterinary Association Clinical Studies Trust Fund (now known as Petsavers).
Presented in part at the British Small Animal Veterinary Association Conference, Birmingham, UK, April 2004.
Address correspondence to Dr. Reid.

made to construct composite scales to measure pain in animals, but although these scales define behaviors that are thought to be indicative of an animal in pain, they have not been shown to fulfill the necessary criteria of a valid, reliable composite measure scale.[6–10] We have previously developed a **composite measure pain scale (CMPS)** for use in dogs with acute pain,[11] adopting the same broad protocols used by Melzack and Torgerson[4] to develop the McGill pain questionnaire. Briefly, a bank of 279 expressions was derived from 69 practicing veterinary surgeons to describe behaviors that were considered to be indicative of pain in dogs. Refinement of the list resulted in 47 expressions (39 behavioral and 8 physiologic signs) remaining. An independent study[12] revealed that physiologic signs were not useful indicators of pain in hospitalized dogs, and these were subsequently removed from the scale. Hence, the prototype CMPS was based on 7 behavioral categories: posture, comfort, vocalization, attention to wound, demeanor, mobility, and response to touch (each containing several expressions describing the dog's behavior). Each expression was precisely defined to ensure consistency between different observers. The validity of the categorization and the assignation of expression within each category were assessed statistically (by use of clustering techniques and the Cronbach α coefficient[13]) and clinically by use of a focus group of 5 qualified veterinary anesthesiologists with extensive experience of animal pain assessment. The tool takes the form of a questionnaire completed by the observer during a prescribed examination procedure, which includes observation of spontaneous behavior and assessment of interactive behavior at rest and during specified movements.

The development of an interval scale was of specific interest to us. Interval level measurement is particularly important in quantitative studies of analgesia because the difference between 2 points on the scale is then immediately interpretable regardless of the position on the scale. To our knowledge, interval scales for measuring pain in animals have not been described to date.

The method by which items in a scale are converted to a number that quantifies appropriately the attribute of interest is known as the scaling model,[14] and this is a vital component in the construction of interval scales. Such models determine how weights are assigned to each item of a scale and how the weights are further combined to produce an overall score. Many scaling models have been developed throughout the history of psychometrics, ranging from very simple to complex theoretical models.[14,15] However, to our knowledge, they have not been used in veterinary medicine.

The assessment of validity is an essential part of the development of a measurement scale. Validity is defined as the effectiveness with which a test or scale measures the property under investigation. A more specific definition states that "validity refers to the appropriateness of inferences from test scores or other forms of assessment."[16] There are several types of validity, of which the most commonly addressed are content, criterion, and construct validity. Content validity focuses on the appropriateness and completeness of the items within the scale and is deemed to be present when the scale items cover all the relevant aspects of the attribute being measured without including any extraneous features. The simplest form of content validity is face validity, which uses expert opinion to establish whether, "on the face of it," the items appear relevant to and encompassing of the test attribute.[17] Criterion validity establishes the effectiveness of the scale's measurement through comparison with a preexisting gold standard method. This can be done by application of the 2 methods simultaneously (concurrent validity) or by testing the ability of the scale to predict future change (predictive validity).[18] However, no gold standard exists for the measurement of pain, although the VAS is often regarded as such within human medicine; Holton et al[3] suggested that, in contrast, the NRS was a more appropriate standard to use in dogs. When a gold standard does not exist, an alternative means of determining validity is to test for construct validity, where a hypothesis is created regarding the scale items, which is then supported or discredited through experiment. Hypotheses used for testing construct validity of pain scales can include the prediction of change in pain scores following the administration of proven analgesics, the ability to discriminate between different severities of surgery, or testing across different patient groups. The process may also involve quantitatively measuring a physical aspect of pain such as mechanical or heat threshold and relating this to the pain score.[10,19–23]

The purpose of the work reported here was, first, to develop the prototype CMPS (**Appendix 1**) into an interval level scale suitable for acute pain assessment in dogs by use of the Thurstone scaling model to calculate weights for each scale item.[24] Second, the validity of the resulting pain tool was to be investigated in dogs within a clinically relevant setting. It is the authors' belief that this is the first report of a validated interval scale to measure clinical pain in an animal species.

Materials and Methods

Application of the Thurstone method of paired comparisons scaling model to the CMPS to create an interval level scale (study 1)—By use of the expressions within each behavioral category (developed previously by our group[11]), pairs of words were created such that each word was compared to each of the remaining expressions and, together with their definitions of behavior, were presented in randomized order to 16 veterinarians. Each person was asked to compare the items in each pair and indicate which word implied higher pain intensity. The estimated probabilities were then transformed into a matrix of z-scores, which indicated the number of SDs between the intensity of each of the items of interest. In view of the fact that the estimates of the differences between the items were likely to be affected by sampling error, the z-scores for each item were averaged over the other items in the scale to reduce the possible error. For ease of use, the weights were transformed to give continuous scores in the range of 0 to 10, on the basis of the principle that the scale would operate such that only 1 item would be chosen from each category.

Construct validity testing of the prototype interval level CMPS (study 2)—Five observers used the CMPS and associated examination procedure to assess pain in 80 dogs (4 groups of 20 dogs). Three groups comprised dogs admitted to the University of Glasgow Veterinary Hospital, whereas the fourth group (control group) consisted of dogs owned by staff at the University of Glasgow. Twenty dogs had undergone orthopedic surgery (orthopedic group), 20 had undergone soft tissue surgery (soft tissue group), and 20 were hospitalized because of medical conditions (medical group). The remaining 20 dogs were determined to be clinically normal (control group). Inclusion criteria placed no restriction in terms of age, breed, surgical procedure, or medical condition, and only dogs that were too aggressive to be easily handled were excluded. The observers, who were all postgraduate students at the University of Glasgow Veterinary Faculty at the time of the study, were qualified veterinarians with experience of veterinary practice. They were not familiar with the dogs included in the study and had had no previous involvement with the development of the CMPS. The objectives of the study, the examination procedure, and use of the measurement scale were explained to each observer.

All examinations were conducted between 12 and 4 PM in the wards of University of Glasgow Veterinary Hospital. When a dog had undergone surgery, the assessments were performed on the day following surgery (ie, from 19 to 29 hours after surgery). Observers were instructed to adhere to the examination procedure defined for the CMPS.[11] Each dog was assessed for as long as was required to identify predominant behaviors and choose the most appropriate descriptors. Typically, the assessments took about 5 minutes to complete, although there was some variation that arose mainly during the assessment of mobility, when some dogs were slower to rise and walk than others. The definitions of the scale items, reasons for hospitalization, and details of surgical procedures undertaken were not available to the observers when making their assessments. All dogs were treated with analgesics according to hospital routine practice, but did not receive any treatment within the hour preceding assessment.

Throughout the study, individual observers assessed the dogs independently of each other, but within the 4-hour observation period (ie, between 12 and 4 PM). If any observer detected signs of an unacceptable degree of pain in a dog, a rescue opioid analgesic was administered (pethidine [3.5 mg/kg, IM]) to that animal and that observer's assessments were excluded from the statistical analyses.

The presence or absence of surgical intervention, the study group (orthopedic, soft tissue, medical, or control), and the perceived severity of pain associated with the surgical procedures or medical conditions were used as the test constructs. Information regarding the perceived severity of pain associated with specific surgical procedures and medical conditions was collected through consultation with 25 veterinarians from University of Glasgow Veterinary Faculty, who had no other involvement in any other stage of the development project. Each was given a list of medical conditions and surgical procedures **(Appendix 2)** and asked to assign one of the following scores: 0, not painful; 1, mild pain; 2, moderate pain; or 3, severe pain. The final score allocated to each condition corresponded to the median value of the 25 collected scores.

Validity and sensitivity assessments (study 3)—Concurrent criterion validity of the scale was assessed by comparing the CMPS scores with scores derived simultaneously from an 11-point NRS applied to dogs undergoing surgery. Construct validity was investigated with 2 constructs. The first construct was the relationship between CMPS scores and time, on the basis of the hypothesis that the scores would change according to the phase of inflammation and stage of healing (ie, scores would be higher on the day of surgery and in the early postoperative period and would decrease with time), and the second was that the CMPS scores would differ between the categories of surgical procedure (orthopedic and soft tissue) and hence the type of tissue trauma involved.

One observer (CMM) used the CMPS and the NRS to assess pain in 77 dogs over a maximum postoperative period of 6 weeks. The dogs had undergone a variety of orthopedic and soft tissue procedures (OR and ST groups, respectively) at the University of Glasgow Veterinary Hospital and had not been included in study 2. The procedures were classified into groups according to the clinical service that undertook the surgery **(Appendix 3)**. As for study 2, no restriction was placed on age, breed, or sex of the dogs assessed, and only those that displayed aggression during their initial examinations were excluded. Analgesia was provided on an individual basis in accordance with standard hospital practice.

The dogs were examined by use of the same procedure as before.[11] The NRS scores were awarded after completing the CMPS protocol. Assessments were undertaken preoperatively (0 hour; n = 77) and at 2 hours (77), 6 hours (75), 24 hours (77), 5 to 7 days (62), 19 to 22 days (57), and 40 to 45 days (45) postoperatively. The initial assessments (0 to 24 hours) were carried out in the University of Glasgow Veterinary Hospital, and the remainder of the assessments (5 to 45 days) were conducted in the dogs' home environments.

Statistical analysis—All statistical analyses were carried out by use of computer software.[a–c] Because of the apparent non-normality in scores in study 2, the relationship between the CMPS score and whether the dog had undergone surgery was investigated by use of a Wilcoxon-Mann-Whitney test. The relationships of study group and perceived pain severity with CMPS scores were investigated by use of the Kruskal-Wallis test. Pairwise comparisons of the median CMPS scores between study groups and pain severity groups were carried out by use of a Wilcoxon-Mann-Whitney test. Values of $P < 0.05$ were considered significant (study 2).

The relationship between NRS and CMPS scores in study 3 was analyzed by use of paired *t* tests. The relationship between time and CMPS scores was examined by use of a repeated-measures ANOVA with Bonferroni pairwise comparisons. Two-sample *t* tests and a repeated-measures ANOVA with Bonferroni pairwise comparisons were used to explore the relationship between the CMPS scores allocated to the OR and ST groups. Values of $P < 0.05$ were considered significant (study 3).

Results

Study 1—Probability matrices were created for each of the 7 behavioral categories **(Table 1)**. Each entry in the matrix indicates the probability that the row item is judged to indicate more of the attribute than the linked column item. The probabilities esti-

Table 1—Representative probability matrix for one of the behavioral categories included in the prototype composite measure pain scale (CMPS); the entries represent the estimated probability of the row item being associated with greater pain intensity than the column item for the "Response to touch" category, which includes 4 items (cry, flinch, growl or guard, and snap). The probabilities indicate the size of the difference between any 2 items.

Response to touch	Cry	Growl or guard	Snap	None
Cry	1.0	0.60	0.27	1.0
Flinch	0.99	0.36	0.18	1.0
Growl or guard	—	—	0.27	1.0
Snap	—	—	—	1.0

— = Not applicable.

mated in the matrix give an indication of the size of the differences between any 2 items. Application of the transformation procedures gave a weight for each scale item (from –2.0 to 2.0; **Table 2**). Further transformation allowed the weights of the scale items to be expressed in the range of 0 to 10. The total pain score was consequently calculated by summing the weights of the items chosen from each category.

Study 2—Eighty dogs of a variety of breeds (age range, 5 months to 15 years) were initially included in the study; 41 dogs were male. During the examination procedure, 3 dogs in the orthopedic group were thought to be in an unacceptable degree of pain. Rescue analgesia was provided for those dogs. Because the data from these dogs were subsequently excluded from all further statistical analyses, 77 dogs were used in the study.

Comparison of the CMPS scores indicated that, for each observer, the dogs that had undergone surgery (n = 37) had higher CMPS scores than those that had not (40). Among all observers, mean ± SD CMPS score was 2.6 ± 1.8 for the dogs that had undergone surgery and 1.2 ± 1.2 for the dogs that had not undergone surgery; the median score was lower for the dogs that had not undergone surgery than the dogs that had (0.9 and 2.4, respectively). Formal analysis revealed a significant (*P* < 0.01) difference between the 2 groups, indicating that the CMPS was sensitive to whether the dogs had undergone surgery.

Table 2—Z–scores and transformed weights for the CMPS behaviors within each of the 7 categories as determined by the Thurstone method of paired comparisons.

Category	Behavior	Z-scores	Transformed weight
Demeanor	Aggressive	0.68	1.22
	Signs of depression	1.37	1.56
	Disinterested	0.77	1.26
	Nervous	0.51	1.13
	Quiet	0.00	0.87
	Content	–1.58	0.08
	Bouncy	–1.74	0.00
Posture	Rigid	0.85	1.20
	Hunched	0.70	1.13
	Normal	–1.55	0.00
Comfort	Restless	1.16	1.17
	Comfortable	–1.16	0.00
Vocalization	Cry	–0.09	0.83
	Groan	0.09	0.92
	Scream	1.74	1.75
	Not vocalizing	–1.74	0.00
Attention to painful area	Chewing	1.24	1.40
	Licking	0.31	0.94
	Ignoring	–1.55	0.00
Mobility	Stiff	0.58	1.17
	Slow	–0.01	0.87
	Lame	1.17	1.46
	Normal	–1.74	0.00
Response to touch	Cry	0.86	1.37
	Flinch	–0.25	0.81
	Snap	0.89	1.38
	Growl	0.36	1.12
	None	–1.86	0.00

Summary statistics of the CMPS scores for each study group for each observer and all observers indicated that the median CMPS scores differed among the 4 condition groups **(Figure 1)**. The orthopedic and soft tissue groups (median score of 3.0 and 2.0, respectively) had higher CMPS scores than the medical and control groups (median score of 1.2 and 0.9, respectively). Also, the maximum observed score was higher for the orthopedic (7.4), soft tissue (6.4), and medical (6.5) groups, compared with that for the control group (4.9). This gives a further indication that the scores allocated by use of the CMPS reflected the differences in the level of pain among groups.

Comparison of the median CMPS scores across the groups by use of a Kruskal-Wallis test revealed significant differences among the 4 groups (χ^2 = 79.98; *P* < 0.001). Pairwise comparisons of the median CMPS scores between the groups indicated that there were significant (*P* < 0.01) differences in the median scores between the 4 study groups.

Most (> 50%) of the surgical procedures were assigned a severity score of 2 (ie, considered to be associated with moderate pain). Of the medical conditions listed, half were thought to cause no pain and only 2 were perceived to be associated with moderate or severe pain. There was high variability in the CMPS scores associated with each severity score. The Spearman rank correlation coefficient between CMPS score and severity (correlation, 0.37) was significantly different from zero, suggesting that the observed CMPS score increased with increasing perceived severity of pain. The median CMPS scores differed among the 4 severity rating groups with the exception of moderate and severe pain (median score was 0.9, 1.8, 2.4, and 2.4 for no pain and mild, moderate, and severe pain, respectively). Comparison of the median CMPS scores across the 4 pain severities by use of the Kruskal-Wallis test revealed that the differences among groups were significant (χ^2 = 57.1; *P* < 0.001). Pairwise compar-

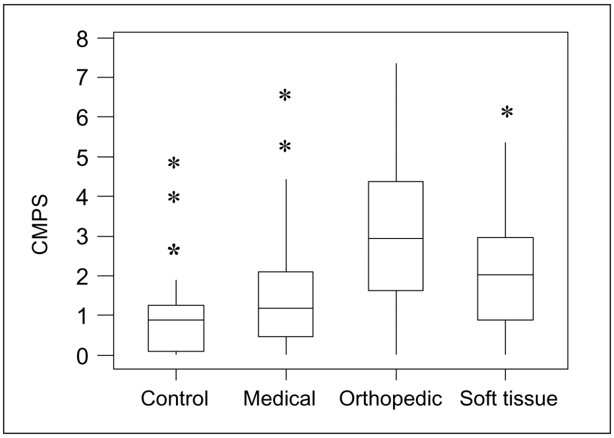

Figure 1—Box and whisker plots of scores obtained by use of a prototype composite measure pain scale (CMPS) in 77 dogs that had undergone orthopedic surgery (n = 17) or soft tissue surgery (20), were hospitalized because of medical conditions (20), or were considered healthy (control; 20). Each dog was assessed by 5 veterinarians; those that had undergone surgery were assessed between 19 and 29 hours later. Each box is the interquartile range (25th or 75th percentile); the horizontal line within each box is the median. Whiskers represent the range. Asterisks represent outliers.

Table 3—Summary statistics for the numerical rating scale (NRS) and the CMPS scores collected in 77 dogs before (0 hour) and at intervals after orthopedic or soft tissue surgery.

Time after surgery	No. of dogs	Method	Mean ± SD	Median	Q1	Q3
0 h	77	NRS	1.05 ± 1.23	1	0	2
		CMPS	1.44 ± 1.13	1.46	0.44	2.33
2 h	77	NRS	2.69 ± 1.38	2	2	4
		CMPS	2.46 ± 1.01	2.33	1.74	3.14
6 h	75	NRS	2.61 ± 1.38	2	1	4
		CMPS	2.18 ± 1.04	2.33	1.68	3.11
24 h	77	NRS	2.20 ± 1.12	2	1	3
		CMPS	1.88 ± 1.12	1.96	0.89	2.79
7 d	62	NRS	1.48 ± 1.04	1.5	1	2
		CMPS	1.52 ± 1.10	1.54	0.08	2.35
21 d	57	NRS	0.95 ± 1.02	1	0	2
		CMPS	1.17 ± 1.15	1.46	0.08	2.35
40 d	45	NRS	0.84 ± 1.09	0	0	1
		CMPS	1.02 ± 1.14	1.46	0	1.54

Q1 = Lower quartile, 25th percentile. Q3 = Upper quartile, 75th percentile.

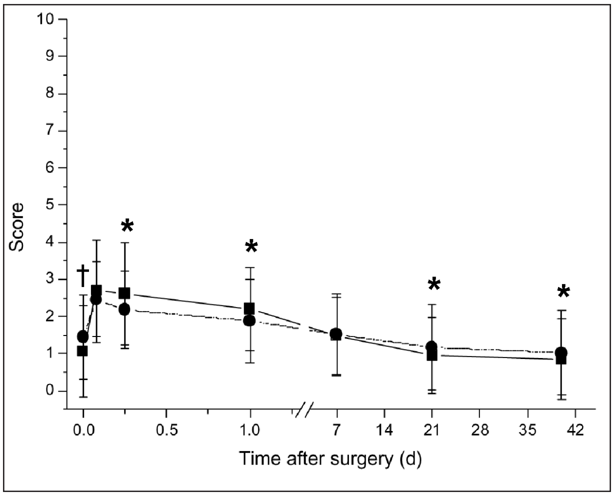

Figure 2—Mean ± SD scores obtained by use of a numeric rating scale (NRS) (squares) and the CMPS (circles) in 77 dogs that underwent orthopedic or soft tissue surgery (assessments were made before [time 0] and after surgery; number of assessments per dog was 4 to 7). *Significant ($P < 0.05$) difference between NRS and CMPS scores. †Significant ($P < 0.01$) difference between NRS and CMPS scores.

isons of the median CMPS scores between the 4 pain severities indicated that the no-pain group had a significantly lower median score than the scores of the other groups. Significant ($P > 0.3$) differences did not exist between the other severity scores.

Study 3—Seventy-seven dogs were enrolled and assessed for 24 hours following surgery; 45 dogs were assessed through 40 to 45 days after surgery. There were various reasons for the attrition in numbers; the owners of 26 dogs were either unavailable at the assessment times or lived more than 1 hour's traveling distance from the hospital, 4 dogs underwent further surgery before the completion of the 6-week assessment period, and 2 dogs were euthanatized. Forty-five (58.4%) dogs were male and 34 breeds were represented; the most common breeds among the 77 dogs were crossbreeds (13.0%), Labrador Retrievers (10.4%), and Golden Retrievers (9.1%). Of the 77 dogs, 42 were allocated to the OR group and 35 to the ST group.

Summary statistics of the pain scores generated by the CMPS and NRS at each time point were calculated, and their means (± SD) over time were compared **(Table 3; Figure 2)**. Most pain scores were distributed toward the lower end of the CMPS (mean, 1.74 ± 1.19; range, 0.00 to 5.87) and NRS ranges (mean, 1.78 ± 1.41; range, 0 to 7). The highest mean pain scores were allocated at 2 hours after surgery (2.46 ± 1.01 for the CMPS and 2.69 ± 1.38 for the NRS).

Results of the paired *t* tests of mean difference (NRS minus CMPS) indicated that the scales were comparable statistically at 2 hours and 7 days after surgery, but the CMPS scores were significantly different from the NRS scores at the other time points (higher at 0 hours and 21 and 40 days and lower at 6 and 24 hours after surgery; **Table 4**). The limits of agreement[25] were –1.68 and 1.77, indicating that the scales were in agreement overall. The scores for both scales followed an equivalent pattern of change over time, peaking at 2 hours and declining to a minimum at 6 weeks after surgery (Figure 2). Sequential changes in the CMPS scores over time were generally smaller than the changes in NRS scores. A highly significant time effect for the CMPS scores was detected by use of a repeated-measures ANOVA (F value, 39.74; $P < 0.001$), confirming that the CMPS was sensitive to changes in pain associated with surgery and healing. The Bonferroni pairwise comparisons indicated that the CMPS scores increased significantly from the preoperative scores to the early postoperative scores (2 and 6 hours; **Table 5**). There was a gradual decline in CMPS scores from 2 hours after surgery, although significant differences did not exist between consecutive time points as a result of minor changes in pain score associated with large SD values. There were no significant differences between the CMPS scores assigned at 7, 21, and 40 days, whereas with the NRS, the decrease in scores between 7 and 40 days was significant. The results indicated that the time intervals associated with significant changes in score were also similar for the 2 scales.

The NRS pain scores supported the hypothesis that orthopedic surgical procedures, compared with soft tissue surgical procedures, are associated with a greater severity of pain (OR group mean score, 2.32 ± 1.30 and ST group mean score, 1.08 ± 1.22), and the

Table 4—Comparison of NRS and CMPS pain scores by use of a paired *t* test of mean differences (NRS minus CMPS) collected in 77 dogs before (0 hour) and at intervals after orthopedic or soft tissue surgery.

Time after surgery	No. of dogs	T value	*P* value	Confidence interval
0 h	77	–4.46	< 0.001*	–0.5679 to –0.217
2 h	77	1.40	0.164	–0.069 to 0.401
6 h	75	3.26	0.002*	0.141 to 0.581
24h	77	3.34	0.001*	0.1218 to 0.4816
7d	62	–0.45	0.652	–0.2233 to 0.1407
21d	57	–2.49	0.016*	–0.4096 to –0.0441
40d	45	–2.13	0.039*	–0.3485 to –0.0097

*Significant ($P < 0.05$) difference between NRS and CMPS pain scores.

Table 5—Results of Bonferroni multiple comparisons for NRS and CMPS scores collected from 41 dogs for which all pain assessments were completed.

Time point	NRS confidence interval	CMPS confidence interval
0-hour time point subtracted from:		
2 h	1.0191 to 2.1029*	0.549 to 1.6126*
6 h	0.9459 to 2.0297*	0.164 to 1.2272*
24 h	0.5069 to 1.5907*	–0.170 to 0.8940
7 d	–0.2492 to 0.8346	–0.591 to 0.4726
21 d	–0.6882 to 0.3955	–0.861 to 0.2026
40 d	–0.8346 to 0.2492	–1.017 to 0.0462
2-hour time point subtracted from:		
6 h	–0.615 to 0.469	–0.917 to 0.146
24 h	–1.054 to 0.030	–1.250 to –0.187*
7 d	–1.810 to –0.726*	–1.672 to –0.608*
21 d	–2.249 to –1.165*	–1.942 to –0.878*
40 d	–2.396 to –1.312*	–2.098 to –1.035*
6-hour time point subtracted from:		
24 h	–0.981 to 0.103	–0.865 to 0.1987
7 d	–1.737 to –0.653*	–1.286 to –0.2228*
21 d	–2.176 to –1.092*	–1.556 to –0.4928*
40 d	–2.322 to –1.239*	–1.713 to –0.6491*
24-hour time point subtracted from:		
7 d	–1.298 to –0.2142*	–0.953 to 0.1104
21 d	–1.737 to –0.6532*	–1.223 to –0.1596*
40 d	–1.883 to –0.7996*	–1.380 to –0.3160*
7-day time point subtracted from:		
21 d	–0.981 to 0.1029	–0.8018 to 0.2618
40 d	–1.127 to –0.0434*	–0.9582 to 0.1055
21-day time point subtracted from:		
40 d	–0.6882 to 0.3955	–0.6882 to 0.3755

*Significant ($P < 0.05$) difference between the mean NRS score at the 2 time points.

2-sample *t* tests (OR minus ST) identified significant differences between the groups over all the scores ($P < 0.001$; 95% **confidence interval [CI]** for difference, 1.16 to 1.70) and at each individual time point (all $P < 0.01$). The greatest differences between the OR and ST group scores were detected at 21 days (95% CI for difference, 1.18 to 1.91), additionally supporting the hypothesis that orthopedic surgical procedures are associated with longer recovery periods. The CMPS scores likewise differentiated between the surgical groups and thus demonstrated good construct validity (OR group mean score, 2.29 ± 0.92 and ST group mean score, 1.01 ± 1.09; T value = 12.75; $P < 0.001$; 95% CI for difference, 1.31 to 1.79; **Figure 3**). Compared with the OR group scores, the ST group scores indicated much greater changes in consecutive pain scores and reached almost baseline values by 21 days after surgery. By comparison, the preoperative OR group scores were higher than the ST group scores and did not differ significantly from the OR group scores at any of the other assessments times. The OR group scores decreased slowly from the early postoperative values (2, 6, and 24 hours), but the difference from the OR group scores in the first 24-hour postoperative period only reached significance at 40 days.

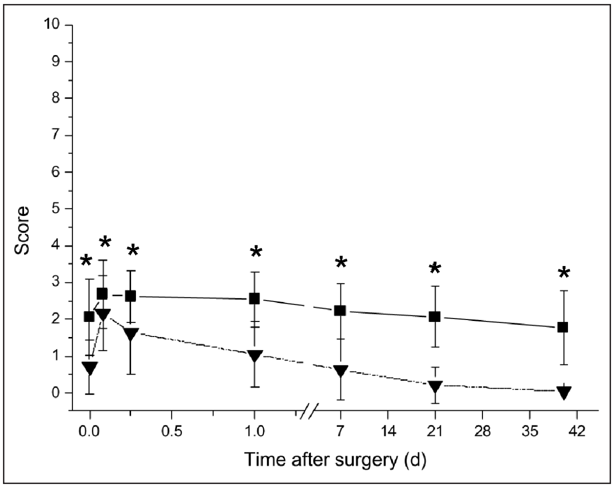

Figure 3—Mean ± SD scores obtained by use of the CMPS in 42 dogs that underwent orthopedic surgery (squares) and 35 dogs that underwent soft tissue surgery (triangles) before (time 0) and after surgery. *Significant ($P < 0.01$) difference between orthopedic and soft tissue groups.

Discussion

The alleviation of pain in animals is an essential aspect of good welfare and an obligation of veterinary practitioners. A prerequisite of this is the ability to recognize and assess pain in animals. Assessment should ideally be structured to quantify pain intensity, identify its sensory quality, and establish the importance of the pain to the well-being of the sufferer. The development of the CMPS was prompted by the need for a valid, reliable, and statistically useful measure of pain in animals. The project of this report was intended to develop the prototype CMPS[11] into an interval level scale and test its validity in clinical situations; results indicated that the scale is sensitive to the different levels of pain induced by orthopedic and soft tissue surgeries or medical conditions and those present in apparently healthy dogs. Therefore, the creation of an interval level scale opens up scope for varied and detailed statistical analyses of pain score results and opportunities for more effective monitoring of acute pain and analgesic efficacy.

It was important to be confident of the validity of the prototype CMPS before further development was undertaken. However, pain measurement, even in humans who can use verbal descriptors, is difficult because pain cannot be directly measured. For this reason, psychometric theory should be applied during the construction of a pain measurement tool to ensure that the resultant scale is valid and actually measures the property of interest. This was done by Melzack and Torgerson[4] during the development of the McGill pain questionnaire.[26] Their work represented a new approach to human pain measurement, not only in terms of the method used but also in terms of aims; until this time, pain had been measured by use of unidimensional scales such as the SDS, NRS, and VAS, which typically only address intensity, and this was the first attempt to reflect the multidimensional nature of pain. There has been some subsequent modification of the original tool,[27] and numerous versions are still in use today, demonstrating the robustness of its structure and development. Therefore, the McGill pain questionnaire was an obvious model on which to base the development of the CMPS.

Psychometric theory dictates that the development of a composite measure scale should involve a number of discrete stages, including careful and thorough item selection and demonstration of face validity and content validity. Well-accepted psychometric theory was applied during the construction of the prototype CMPS; inherent validity was implied, and face validity and content validity were subsequently determined, indicating that this was an acceptable scale on which to apply the Thurstone scaling model.

An important difference between the CMPS and the McGill pain questionnaire is that the former relies on the recognition and interpretation of behavioral signs by an independent observer rather than on self-reported data. The clinical implication is that interobserver variability must be investigated and accounted for. Holton et al[3] investigated the reliability of the SDS, NRS, and VAS for the assessment of pain in dogs, and their findings indicated that when 3 and 4 veterinarians simultaneously scored pain on the day of and the day following surgery, respectively, the interobserver variability was unacceptable. A further goal of the CMPS development was to produce an observational scale that would produce consistent, valid results across observers.

The problem of interobserver variability has been addressed during the development of tools to monitor other functions such as the level of consciousness, notably in the widely used Glasgow coma scale.[28] This scale focuses on 3 different aspects of behavioral response. The universality of the scale depends on identifying responses that can be clearly defined and accurately graded according to a rank order that indicates the degree of dysfunction. According to Teasdale and Jennett,[28] for a scale to be generally accepted as universal, it must be practical to use in a wide range of locations and by staff without special training. For this reason, it is important to define behavioral descriptors precisely so that the observer is left in no doubt as to the interpretation. Because our intent was to develop a universal pain scale for dogs, suitable for use by lay people as well as health professionals, it was deemed necessary to include a list of definitions for all expressions used in the scale. By doing this, the need for training is reduced; the scale is presented so that the method for completion is clearly understood, and the detailed definitions of behavior prevent confusion when descriptors are chosen.

As stated, an essential part of the initial development was to ensure that the items chosen for inclusion in the CMPS had content validity. The initial item list included several physiologic signs that have previously been considered to be indicative of acute pain. Physiologic measures have been incorporated together with behavioral signs in composite tools designed for use in young children[29] and dogs[10]; however, despite the fact that physiologic measures are objective and precise, Anand and Hickey[30] determined that they are nonspecific for pain in neonates, and similar findings in dogs were reported by Holton et al.[12] On the basis of this, physiologic signs were removed from the CMPS. In contrast, the University of Melbourne pain scale described by Firth and Haldane[10] includes heart rate and respiratory rate. That composite pain scale has some similarities to the CMPS insofar as it was developed to evaluate postoperative pain in dogs and includes behavioral signs such as response to palpation, activity, mental status, posture, and vocalization; however, details of how item selection was undertaken and proof of content validity were not provided. Another important difference between the CMPS and the University of Melbourne pain scale is that the latter does not include a scaling component, which underpins the validity of a scale.

Regardless of the precision with which the CMPS was developed, the interpretation of behavior-based scales is complicated by the fact that the behavioral responses are themselves subject to change. Lilley et al[31] determined that during the first 18 months of life, facial expressions have important implications for the understanding and assessment of human pain, but that these expressions change in an age-related manner. Furthermore, Hamers et al[32] identified external factors (such as environment) that can further modify human behavioral expressions of pain. Lack of specificity of behavioral responses to pain also exists; in studies by Barr[33] and Chambers et al,[34] the same response could be induced by physiologic states such as fear, stress, and hunger. In veterinary medicine, the recognition of different manifestations of pain is further complicated by major species and breed differences; hence, the CMPS was developed to be used in 1 species and tested across a range of breeds.

Empirical verification of the validity of the interval CMPS was a fundamental part of the present study. The first construct investigated, whether the scores differed between dogs that had undergone surgery and those that had not, provided support for its validity. To explore the validity of the scale further and demonstrate a fine level of differentiation, the differences in pain scores associated with the 4 study groups of dogs were investigated, and the hypothesis that the orthopedic group would have the most pain, followed in decreasing order by the soft tissue group and medical group, and that the control group would have the lowest pain scores was tested. The median CMPS scores supported this and provided further evidence that the CMPS is valid when used in a clinical setting.

The data acquired from the study investigating the relationship between the CMPS score and the perceived intensity of pain for each treatment severity were not as conclusive as for the other 2 constructs. There were significant differences in the median CMPS scores between the no-pain group and each of the other groups, but not among the other individual groups. In this type of testing, it is possible that the construct itself is not robust because it is based on subjective judgments of the degree of pain expected to be associated with certain surgical procedures. Results of surveys by Dohoo and Dohoo[35] and Capner et al[36] illustrated that a wide variety of opinions exist regarding the severity of postoperative pain, whereas few studies have specifically examined the accuracy of these opinions. Indeed, Fox et al[37] investigated pain associated with ovariohysterectomy in dogs and reported the procedure to be more painful than commonly regarded. The relatively poor correlation between CMPS scores and mild, moderate, and severe pain severity scores could also be explained in part by the use of analgesia in the patients. Ethically, it was necessary to provide appropriate analgesia, and this was carried out independently of our study and in accordance with routine hospital practice. As a consequence, the level of pain in each dog was modified because analgesia was tailored to individual need and governed by the invasiveness of the procedure and the severity of the condition. Hence, it is probable that dogs that had undergone surgeries perceived to induce severe pain had their pain controlled to the extent that they appeared comfortable when assessed and were awarded low CMPS scores.

The constructs of the presence or absence of surgical intervention, study group, and perceived pain severity were highly interrelated; therefore, further validity testing was undertaken involving the additional, unrelated construct of time and also by introducing concurrent criterion validity. Time is a well-recognized construct used to support the validity of pain scales in both human and veterinary medicine. Early studies[38,39] in dogs used time as a single construct to investigate the performance of the VAS, whereas validation of a postoperative pain measure for parents involved comparing scores over 2 days following surgery in children and, at the same time, evaluating scores against each child's self-report of pain.[21] The investigation of the validity of the COMFORT scale, designed to assess pain in children 3 years old or less, was carried out during time intervals until 36 hours after surgery and, in common with the study of this report, used a proxy scale as a second method of comparison.[29]

A notable difference between the present and previous studies is the duration of postoperative assessments. In most previous reports, the assessment period did not proceed beyond 24 hours after surgery,[40–43] and indeed several were considerably shorter, lasting no more than a few hours after tracheal extubation.[37–39]

It is generally accepted that surgery-associated pain is most severe in the early postoperative period[9] and then decreases with time as inflammation subsides and healing occurs. Clinical experiences indicate gradual amelioration of discomfort over approximately 10 to 14 days after soft tissue surgeries and 4 to 6 weeks after orthopedic surgeries. The CMPS scores collected in the present study reflected this observation, although significant differences did not exist between consecutive postoperative scores. This may reflect a gradual lessening of pain in a sample with large individual variation in pain level, or alternatively, it could arise because of a lack of responsiveness of the scale. However, results of several studies[29,38,39,44,45] in humans and other animals have indicated that low levels of postoperative pain are not unusual, and thus significant changes in score with healing will be less obvious. De Kock et al[44] reported that among humans undergoing surgery to treat rectal adenocarcinoma, most people self-reported scores of 40 and below via a VAS, even during coughing; in a similar investigation by Ilkjaer et al[45] of women undergoing ovariohysterectomy, the VAS scores ranged from a median of 30 or below at rest to approximately 60 during coughing. When the VAS was used as an observational scale to assess children undergoing major abdominal and thoracic surgery, the VAS score never exceeded 4 of 10 for 32% of infants.[29] Within veterinary medicine, Reid and Nolan[38] and Nolan and Reid[39] reported VAS scores of 30 and below following a variety of soft tissue and orthopedic procedures in dogs. Adequate perioperative analgesia, which was provided to all the dogs in the present study, would mask the effects of time on pain scores because the very premise of perioperative analgesia is to ameliorate postoperative pain.

The NRS scores had slightly greater changes over time, compared with the CMPS scores, which raises the possibility that the CMPS was comparatively less responsive. Responsiveness is defined as the ability of a tool to measure a change in the attribute being investigated.[46] The level of responsiveness required from a scale depends on the variation in the level of the attribute within the population and, possibly more importantly, on what constitutes the minimal clinically important difference (ie, the amount of change that is considered important to the patient).[47,48] Barr[33] maintained that it is possible to estimate the responsiveness of a scale by applying an intervention of known efficacy and measuring the magnitude of change. For a pain scale, an analgesic drug of predictable potency could be used as the intervention, and the difference that it produced in pain score would then be compared to the within-subject variation to give the index of responsiveness. Results of a large clinical trial in humans[49] indicated that a reduction of 30% or 2 points on an 11-point NRS was deemed to be meaningful and that this was consistent across a range of painful conditions; however, with less intense pain, a 0.5-point reduction in the NRS score can correspond with a judgment that the pain status is much improved. Testing the responsiveness of the CMPS in dogs would be best carried out in a study designed to assess the response to analgesic treatment.

The construct relating severity of surgical procedure to CMPS pain scores was revisited in the latter part of the present study, and the difference in the duration of the recovery period provided additional discrimination. There was an unequivocal difference between the pain scores for the dogs of the OR and ST groups at each time point and in their respective recovery times.

Concurrent validity testing can be considered to be more robust than construct validity assessment because uncertainty regarding the accuracy of the construct is avoided. However, it does depend on choosing an appropriate gold standard with which comparisons are drawn. The VAS has been considered to be the gold standard pain measure in humans[50] and has been widely used in veterinary pain assessment.[10,38,40,51,52] Its purported benefits include a high degree of sensitivity (arising from the continuous nature of the scale) and interval level measurement. The level of sensitivity required for effective pain measurement in humans was determined by Jensen et al,[53] who compared the information provided by 3 NRS scales of differing levels of detail. It was concluded that little discriminatory power was lost by recoding the 101-point scale as either 11- or 21-point scales, suggesting that an 11-point NRS would be a sufficiently sensitive measure of pain in animals, for which the number of identifiable pain levels would likely be fewer. One of the main disadvantages of the VAS is that it is a written scale; errors of as much as ± 20 on a 100-point scale can be introduced by the visual acuity of the user and the format in which the scale is presented.[54–56] These factors may have accounted for the poorer reliability associated with the VAS in the interobserver reliability study of Holton et al,[3] from which the authors concluded the NRS to be the most appropriate unidimensional pain scale for use in dogs.

One of the potential difficulties associated with concurrent validation is that experimental protocol may encourage similarities to appear between the tools. In the present study, the dogs were examined in accordance with the CMPS protocol before the NRS score was assigned, and as such, the NRS score was influenced by observation of the same composite behaviors that contributed to the CMPS score. Controlled behavioral assessments influence self-reports of chronic pain in humans, improving the correlation between the unidimensional pain score and observed pain behaviors.[57] Randomizing the order in which the scales were completed could have addressed this, but it was considered that this would have introduced unacceptable discrepancies because a proportion of NRS scores would be assigned prior to the behavioral assessment and thus would be based entirely on the observation of spontaneous behavior. Videotaped data of canine behavior collected following ovariohysterectomy revealed that pain modifies both spontaneous and interactive behavior, and accurate pain assessment must take account of both.[37]

Compared with the NRS scores, the overall variation in scores was less for the CMPS, being most marked at 2 and 6 hours postoperatively. The CMPS was purposely designed to minimize the subjective interpretation of pain behaviors with the intent of improving interobserver reliability. Comparatively, the NRS is susceptible to many influences, which in some instances could be considered advantageous. For example, if a dog is vocalizing because of being hospitalized rather than because of pain, the NRS score can be based on other signs of pain (having judged vocalization to be not useful in that particular situation). Nevertheless, it is possible that the significantly greater variance in the NRS scores in the immediate postoperative period arose through subjective assumptions used to compensate for effects of sedation or from preconceived expectations of the severity of pain associated with the surgery. A further explanation for this may be that the NRS does not show interval level measurement in this situation, on the basis of the observation that the greatest SD values were associated with the highest mean pain scores. Farrar et al[49] reported that people generally used the NRS in such a way that the difference in pain between points at the lower end of the scale was smaller than the difference in pain at the middle or upper extreme of the scale; therefore, the NRS did not function as an interval scale.

Although the development of the CMPS was modeled on the methodology of the McGill pain questionnaire, it was considerably augmented by the first use in veterinary medicine of a scaling model designed to provide interval level measurement. The concept of equal intervals between consecutive points on the scale is crucial to the measurement of pain because pain intensity is assumed to follow an underlying continuous distribution. Measuring such an attribute via nominal or ordinal level scales results in loss of information because the continuous measurement is missing and the difference between categories cannot be guaranteed to be constant. Therefore, such scales do not adequately reflect the experience of the patient or differences between groups of patients; in addition, they are limited in the ways that the data can be treated statistically. In contrast, interval level measurement allows the difference between any response and a constant to be known.

Two main types of scaling model are commonly used in psychometrics to provide interval level measurement: the theoretical (or subjective estimate) model and the empirical (or discriminant) model. A theoretical model is based on the judgment of the investigator and accepted theory. An empirical model is based on information gathered during an investigation of the relationships between items in the scale, which then allows a weight for each item to be derived accordingly. In the more recent pain measurement scales developed for use in animals, no information is provided on the derivation of the item weights.[9,10,40] Therefore, it must be assumed that the authors used subjective estimation techniques that were based on their best estimate of the most appropriate weights. Although Firth and Haldane[10] confirmed that scores derived from 3 of the behavior categories (mental activity, posture, and vocalization) changed according to the severity of pain, no differentiation was provided by activity scores. If a pain measurement scale is to accurately reflect an animal's pain experience, information relating behavioral change to pain intensity must be included in the weighting system, yet the true relative importance of behaviors is seldom explored in subjective estimation techniques, and this represents a major disadvantage.

An empirical scaling approach has been adopted in the CMPS; the approach was based on the Thurstone model of matched pairs derived from the classical law of comparative judgment proposed by Thurstone in 1928.[24] When the Thurstone method of paired comparisons is applied, the measurement scale should be structured as small groups of items, each relating to the same attribute but associated with differing levels. The respondent is then required to choose the item in each group that most appropriately represents the level of attribute observed. Because these assumptions hold for the CMPS in general, this model was deemed appropriate to its design. Researchers in psychology have determined that scores produced by use of this model possess the properties of interval level measurement.[14]

It is the application of these statistically derived weights to the word descriptors that creates interval level measurement and distinguishes the CMPS from other composite measure scales.[9,10] In these scales, descriptors have been ranked and allocated a numerical value that reflects their rank position. The composite score is thus generated from the sum of the ranks. Although this appears to give a continuous measurement, it only gives ordinal information, which restricts the type of analysis that can be used on the scores (nonparametric vs more powerful parametric analysis). In contrast, the weights allocated in the CMPS allow the descriptors to be positioned along the scale continuum at intervals that indicate the amount of pain represented by the word. The CMPS contains descriptors that are positive and negative indicators of pain, and this is reflected by the weights so that the descriptors "happy and bouncy" and "depressed" are given low and high weights accordingly and positioned at opposite ends of the continuum. Furthermore, this type of scaling means that descriptors can be used to describe quite different behaviors, even if these behaviors imply similar levels of pain (eg, "happy and content" and "happy and bouncy"). This avoids the need for exhaustive lists of adjectives and allows the scale to be concise and practical.

By use of this scaling model, it would be possible to reassign weights to give scales that are specific to certain types of pain if experiments revealed that certain types of behavior were more characteristic of pain arising from certain sources (eg, soft tissue vs orthopedic injury). However, the CMPS was designed to be universally suitable and hence was tested and determined to be sensitive across a range of painful conditions.

An important feature of interval scales is that the zero score is arbitrary, in contrast to ratio scales, in which the zero score reflects an absence of the attribute. When pain is assessed in animals, the patient cannot convey the sensory and emotional experience by any means other than its behavior. Therefore, only levels of pain that cause the animals to alter their behavior in some way can be conveyed to the observer. In humans, some discordance has been shown to exist between patients' self-reports of pain and pain assessed via behavioral observation.[58] This suggests that the patients' behavior did not fully represent their pain, and it is possible for a patient to be experiencing a degree of pain that would be reported on a scale, but that may not be sufficient to alter behavior. A score of zero that is based on behavior cannot therefore be assumed to indicate the complete absence of pain.

Our data indicate that the methodology used in the design of the CMPS conferred face and content validity, which provided a sound base for further development. With the increasing importance of an evidence-based approach to medicine comes the need to measure outcome, and that is achieved most effectively through the use of scales that provided continuous, interval level measurement. To our knowledge, this has not existed in veterinary medicine to date. Composite behavioral scales have the advantage that they guide the observer toward specific types of behavior and can therefore improve the reliability of observation.[28] In addition, their structure is suitable for the creation of interval level scales by the application of the Thurstone model of matched pairs, which allows word descriptors to be placed at meaningful intervals along the continuum of the scale. Validity testing is also fundamental to effective measurement—a pain scale must be shown to measure pain. The validity of the CMPS to measure pain within a clinical situation has been supported by a number of methods of testing. Nonetheless, the difficulty of establishing validity in a scale that measures such an intangible entity as pain means that further testing is still required. It would be advantageous to find an objective proxy measure of a related factor such as pain hypersensitivity for comparison.[59] A clinical investigation, such as presented here, inevitably has many possible sources of variation that may influence the results and their interpretation; future studies may benefit from being carried out in controlled conditions and involving, for example, groups of dogs of the same breed that were undergoing the same surgical procedure conducted by the same surgeon and exposed to the same analgesic protocol. Although further work is required to improve the understanding of how the CMPS performs and modification may be required before it can be used universally, results of initial studies are encouraging and suggest that the CMPS has the potential to become a valuable and versatile tool.

a. SAS, version 12.0 for Windows, SAS Institute Inc, Cary, NC.
b. MINITAB, version 13, Minitab Inc, Microsoft Corp, Coventry, UK.
c. Origin 6 software, Microcal Software Inc, Northampton, Mass.

References

1. Merskey H, Bogduk N. Chapter number or section title. In: *Classification of chronic pain: descriptions of chronic pain syndromes and definitions of pain terms*. Seattle: ISAP Press, 1994;209–214.
2. Zimmerman M. Behavioural investigations of pain in animals. In: Duncan IJH, Molony V, eds. *Assessing pain in farm animals*. Luxembourg: European Communities, 1986;16–27.
3. Holton L, Scott EM, Nolan AM, et al. Comparison of three methods used for assessment of pain in dogs. *J Am Vet Med Assoc* 1998;212:61–66.
4. Melzack R, Torgerson WS. On the language of pain. *Anesthesiology* 1971;34:50–59.
5. Coste J, Fermanian J, Venot A. Methodological and statistical problems in the construction of composite measurement scales: a survey of six medical and epidemiological journals. *Stat Med* 1995;14:331–345.
6. Morton DB, Griffiths PH. Guidelines on the recognition of pain, distress and discomfort in experimental animals and an hypothesis for assessment. *Vet Rec* 1985;116:431–436.
7. Sandford J, Ewbank R, Molony V, et al. Guidelines for the recognition and assessment of pain in animals. *Vet Rec* 1986;118:334–338.
8. Dodman N, Clarke GH, Court MH, et al. Epidural opioid administration for post operative pain relief in the dog. In: Short CE,

Van Pozna A, eds. *Animal pain*. New York: Churchill Livingstone Inc, 1992;274–277.

9. Hellyer PW, Gaynor JS. Acute postsurgical pain in dogs and cats. *Compend Contin Educ Pract Vet* 1998;20:140–153.

10. Firth AM, Haldane SL. Development of a scale to evaluate postoperative pain in dogs. *J Am Vet Med Assoc* 1999; 214:651–659.

11. Holton L, Reid J, Scott EM, et al. Development of a behaviour-based scale to measure acute pain in dogs. *Vet Rec* 2001;148:525–531.

12. Holton L, Scott EM, Nolan AM, et al. Relationship between physiological factors and clinical pain in dogs scored using a numerical rating scale. *J Small Anim Pract* 1998;39:469–474.

13. Cronbach LJ. Coefficient Alpha and the internal structure of tests. *Psychometrica* 1951;16:297–334.

14. Nunnally JC, Bernstein IH. *Psychometric theory*. 3rd ed. New York: McGraw-Hill Book Co, 1994.

15. Torgerson WS. *Theory and methods of scaling*. New York: John Wiley & Sons, 1958.

16. Ghiselli EE, Campbell JP, Zedeck S. *Measurement theory for the behavioural sciences*. Oxford, England: WH Freeman, 1981.

17. Frayers PM, Hand DJ. Causal variables, indicator variables and measurement scales: an example from quality of life. *J R Stat Soc [Ser A]* 2002;165:233–261.

18. Bowling A. Theory of measurement. In: *Measuring health: a review of quality of life measurement scales*. Buckingham, UK: Open University Press, 1997;12–22.

19. Chapman CR, Casey KL, Dubner R, et al. Pain measurement: an overview. *Pain* 1985;22:1–31.

20. Price DD, Bush FM, Long S, et al. A comparison of pain measurement characteristics of mechanical visual analogue and simple numerical rating scales. *Pain* 1994;56:217–226.

21. Chambers CT, Reid GJ, McGrath PJ, et al. Development and preliminary validation of a postoperative pain measure for parents. *Pain* 1996;68:307–313.

22. Melzack R, Katz J. Pain measurement in persons in pain. In: Melzack R, Wall P, eds. *Textbook of pain*. Edinburgh: Churchill Livingstone, 1999;409–425.

23. Bodian CA, Freedman G, Hossain S, et al. The visual analog scale for pain: clinical significance in postoperative patients. *Anesthesiology* 2001;95:1356–1361.

24. Thurstone LL. Law of comparative judgment. *Am J Sociol* 1928;33:529–554.

25. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307–310.

26. Melzack R. The McGill Pain Questionnaire: major properties and scoring methods. *Pain* 1975;1:277–299.

27. Clark WC, Fletcher D, Janal M, et al. Hierarchial clustering of pain and emotion descriptors: towards a revision of the McGill pain questionnaire. *Adv Pain Res Ther* 1995;22:319–330.

28. Teasdale G, Jennett B. Assessment of coma impaired consciousness: a practical scale. *Lancet* 1974;2:81–84.

29. van Dijk M, de Boer JB, Koot HM, et al. The reliability and validity of the COMFORT scale as a postoperative pain instrument in 0 to 3-year-old infants. *Pain* 2000;84:367–377.

30. Anand KJ, Hickey PR. Pain and its effects in the human neonate and fetus. *N Engl J Med* 1987;317:1321–1329.

31. Lilley CM, Craig KD, Grunau RE. The expression of pain in infants and toddlers: developmental changes in facial action. *Pain* 1997;72:161–170.

32. Hamers JP, Abu-Saad HH, van den Hout MA, et al. The influence of children's vocal expressions, age, medical diagnosis and information obtained from parents on nurses' pain assessments and decisions regarding interventions. *Pain* 1996;65:53–61.

33. Barr RG. Reflections on measuring pain in infants: dissociation in responsive systems and "honest signaling." *Arch Dis Child* 1998;79:152–156.

34. Chambers CT, Giesbrecht K, Craig KD, et al. A comparison of faces scales for the measurement of pediatric pain: children's and parents' ratings. *Pain* 1999;83:25–35.

35. Dohoo SE, Dohoo IR. Factors influencing the postoperative use of analgesics in dogs and cats by Canadian veterinarians. *Can Vet J* 1996;37:552–556.

36. Capner CA, Lascelles BD, Waterman-Pearson AE. Current British veterinary attitudes to perioperative analgesia for dogs. *Vet Rec* 1999;145:95–99.

37. Fox SM, Mellor DJ, Stafford KJ, et al. The effects of ovariohysterectomy plus different combinations of halothane anaesthesia and butorphanol analgesia on behaviour in the bitch. *Res Vet Sci* 2000;68:265–274.

38. Reid J, Nolan AM. A comparison of the postoperative analgesic and sedative effects of flunixin and papaveretum in the dog. *J Small Anim Pract* 1991;32:603–608.

39. Nolan A, Reid J. Comparison of the postoperative analgesic and sedative effects of carprofen and papaveretum in the dog. *Vet Rec* 1993;133:240–242.

40. Conzemius MG, Hill CM, Sammarco JL, et al. Correlation between subjective and objective measures used to determine severity of postoperative pain in dogs. *J Am Vet Med Assoc* 1997;210:1619–1622.

41. Hardie EM, Hansen B, Carroll GS. Behaviour after ovariohysterectomy in the dog. *Appl Anim Behav Sci* 1997;51:111–128.

42. Lascelles BD, Cripps PJ, Jones A, et al. Post-operative central hypersensitivity and pain: the pre-emptive value of pethidine for ovariohysterectomy. *Pain* 1997;73:461–471.

43. Kyles AE, Hardie EM, Hansen BD, et al. Comparison of transdermal fentanyl and intramuscular oxymorphone on post-operative behaviour after ovariohysterectomy in dogs. *Res Vet Sci* 1998;65:245–251.

44. De Kock M, Lavand'homme P, Waterloos H. 'Balanced analgesia' in the perioperative period: is there a place for ketamine? *Pain* 2001;92:373–380.

45. Ilkjaer S, Bach LF, Nielsen PA, et al. Effect of preoperative oral dextromethorphan on immediate and late postoperative pain and hyperalgesia after total abdominal hysterectomy. *Pain* 2000;86:19–24.

46. Wright JG, Young NL. A comparison of different indices of responsiveness. *J Clin Epidemiol* 1997;50:239–246.

47. Prasad K. The Glasgow Coma Scale: a critical appraisal of its clinimetric properties. *J Clin Epidemiol* 1996;49:755–763.

48. Bellamy N, Campbell J, Haraoui B, et al. Dimensionality and clinical importance of pain and disability in hand osteoarthritis: development of the Australian/Canadian (AUSCAN) Osteoarthritis Hand Index. *Osteoarthritis Cartilage* 2002;10:855–862.

49. Farrar JT, Portenoy RK, Berlin JA, et al. Defining the clinically important difference in pain outcome measures. *Pain* 2000; 88:287–294.

50. De Williams AC, Davies HT, Chadury Y. Simple pain rating scales hide complex idiosyncratic meanings. *Pain* 2000;85:457–463.

51. Lascelles BD, Butterworth SJ, Waterman AE. Postoperative analgesic and sedative effects of carprofen and pethidine in dogs. *Vet Rec* 1994;134:187–191.

52. Sammarco JL, Conzemius MG, Perkowski SZ, et al. Postoperative analgesia for stifle surgery: a comparison of intra-articular bupivacaine, morphine, or saline. *Vet Surg* 1996;25:59–69.

53. Jensen MP, Turner JA, Romano JM. What is the maximum number of levels needed in pain intensity measurement? *Pain* 1994;58:387–392.

54. Dixon JS, Bird HA. Reproducibility along a 10 cm vertical visual analogue scale. *Ann Rheum Dis* 1981;40:87–89.

55. Sriwatanakul K, Kelvie W, Lasagna L, et al. Studies with different types of visual analog scales for measurement of pain. *Clin Pharmacol Ther* 1983;34:234–239.

56. Bodian CA, Freedman G, Hossain S, et al. The visual analog scale for pain: clinical significance in postoperative patients. *Anesthesiology* 2001;95:1356–1361.

57. Labus JS, Keefe FJ, Jensen MP. Self-reports of pain intensity and direct observations of pain behavior: when are they correlated? *Pain* 2003;102:109–124.

58. Teske K, Daut RL, Cleeland CS. Relationships between nurses' observation and patients self-report of pain. *Pain* 1983;16:289–296.

59. Dolan S, Kelly JG, Monteiro AM, et al. Up-regulation of metabotropic glutamate receptor subtypes 3 and 5 in spinal cord in a clinical model of persistent inflammation and hyperalgesia. *Pain* 2003;106:501–512.

Appendix 1

Prototype composite measure pain scale questionnaire. From their observations, evaluators were requested to choose one of the descriptors as the answer to the following questions regarding a dog's spontaneous behavior. After completion of the questionnaire, evaluators were requested to make an overall assessment of the pain that they considered the dog to have by use of a numeric rating scale; a number from 0 to 10 (0 indicated that the dog was in no pain and 10 indicated that the dog's pain could not be worse) was assigned to each dog.

Variable	Question	Descriptor
Posture	Does the dog seem…	Rigid? Hunched or tense? To have normal posture?
Comfort	Does the dog seem…	Restless? Comfortable?
Vocalization	If the dog is vocalizing, is it…	Crying or whimpering? Groaning? Screaming? Not vocalizing or none of these?
Attention to wound area	Is the dog…	Chewing the wound? Licking, looking at, or rubbing the wound? Ignoring the wound?
Demeanor	Having interacted with the dog, does the dog seem to be…	Aggressive? Depressed? Disinterested? Nervous, anxious, or fearful? Quiet or indifferent? Happy and content? Happy and bouncy?
Mobility*	After walking each dog for a short distance (if possible), did the dog seem to be…	Stiff? Slow or reluctant to rise or sit? Lame? Able to move with normal gait? Assessment not carried out?
Response to touch	When gentle, even pressure was applied to the area approximately 2 inches around the surgical wound (or near the area of the wound if it is inaccessible), did the dog…	Cry? Flinch? Snap? Growl or guard wound? Have no adverse response to touch?

*In some instances, this assessment is not possible because of the type of surgery.

Appendices are continued on the next page

Appendix 2

Orthopedic and soft tissue surgical procedures and medical conditions presented to 25 veterinarians to allow assessment of the severity of pain associated with each.

Orthopedic procedures	Soft tissue procedures	Medical conditions
Amputation of digit	Anal gland removal	Acute moist dermatitis
Arthrotomy—carpus	Anal furunculosis treatment	Acute otitis externa
Arthrotomy—elbow joint	Biopsy—gastrointestinal tract	Acute pancreatitis
Arthrotomy—shoulder joint	Biopsy—liver	Hypoadrenocorticism
Arthrotomy—stifle joint	Biopsy—soft tissue mass	Chronic nephritis
Carpal arthrodesis	Castration	Chronic otitis externa
Cruciate ligament repair	Cataract removal	Hyperadrenocorticism
Forelimb amputation	Cryosurgery	Diabetes mellitus
Fracture repair—plate	Cystotomy	Dilated cardiomyopathy
Fracture repair—pin	Diaphragmatic hernia repair	Endocardiosis
Hind limb amputation	Entropion correction	Focal erosive gastritis
Joint flush	Exploratory thoracotomy	Hepatic failure
Laminectomy	Enucleation	Lymphoma
Major tooth extraction	Implant nasal drain	Osteosarcoma
Mandibulectomy	Lateral wall resection	Pyrexia
Minor tooth extraction	Lung lobectomy	Vomiting and diarrhea
Patella—lateral capsular overlap	Ovariohysterectomy	
Repair of luxated hip joint	Perineal hernia repair	
Total hip replacement	Prostatic cyst removal	
Tibial crest transplant	Removal of intestinal foreign body	
Triple pelvic osteotomy	Soft tissue mass resection (approx 3 cm)	
Ventral slot	Soft tissue mass resection (approx 10 cm)	
Vertebral distraction	Soft palate resection	
	Suture pad	
	Total ear canal ablation	
	Tonsillectomy	
	Urethrotomy	

Appendix 3

Orthopedic and soft tissue procedures (classified in terms of the clinical service that undertook the surgery) performed on 77 dogs.

Orthopedic procedures	Soft tissue procedures
Achilles rupture repair (n = 2)	Castration (n = 3)
Amputation of digit (n = 1)	Cryptorchid castration (n = 1)
Arthroscopy (n = 4)	Cystotomy (n = 1)
Arthrotomy (n = 2)	Ectopic ureter implantation (n = 3)[a]
Cruciate ligament repair (n = 5)	Enucleation (n = 1)
Fractured calcaneous repair (n = 1)	Excise mass and skin flap (n = 1)
Fractured metacarpal bone repair (n = 1)	Exploratory laparotomy (n = 2)
Fractured tibia and fibula repair—external fixator (n = 1)	Gastropexy (n = 1)
Fractured pelvis repair (n = 1)	Correction of intussusception (n = 1)
Fracture revision (n = 1)	Mandibulectomy (n = 2)
Forelimb amputation (n = 1)	Ovariohysterectomy (n = 6)
Hemilaminectomy (n = 1)	Ovariohysterectomy and remove vaginal mass (n = 1)
Hind limb amputation (n = 1)	Patent ductus arteriosus ligation (n = 2)
Implant removal (n = 2)	Perianal mass excision and castration (n = 1)
Radial ostectomy (n = 1)	Perineal hernia correction (n = 1)[a]
Tarsal arthrodesis (n = 3)	Perineal hernia correction and castration (n = 1)[a]
Tibial crest transposition (n = 2)	Perineal mass excision (n = 1)
Total hip replacement (n = 4)	Pinnectomy (n = 1)
Tibial plateau leveling osteotomy (n = 4)	Portosystemic shunt ligation (n = 1)
Triple pelvic osteotomy (n = 1)	Total ear canal ablation and bulla osteotomy (n = 3)[a]
Trochleoplasty (n = 1)	Thyroidectomy (n = 1)
Ulnar osteotomy (n = 1)	
Vertebral stabilization (n = 1)	

[a]Three dogs underwent bilateral procedures on separate occasions and were assessed after each surgery. These procedures were bilateral total ear canal ablation and bulla osteotomy, perineal hernia repair (1 surgery also involving castration), and bilateral ureter implantation.