TITLE: Inter-rater and inter-device reliability of mechanical thresholds measurement with the Electronic von Frey Anaesthesiometer and the SMALGO in healthy cats

AUTHORS: Adami, C; Lardone, E; Monticelli, P

JOURNAL TITLE: Journal of Feline Medicine and Surgery

PUBLICATION DATE: 22 November 2018 (online)

PUBLISHER: SAGE Publications

DOI: https://doi.org/10.1177%2F1098612X18813426

Original article

Inter-rater and inter-device reliability of mechanical thresholds measurement with the Electronic von Frey Anaesthesiometer and the SMALGO in healthy cats

Chiara Adami[1], Elena Lardone[2] and Paolo Monticelli[1]

[1]Department of Clinical Sciences and Services, Royal Veterinary College, University of London, UK

[2]Department of Veterinary Science, University of Turin, Italy

Corresponding author:

Chiara Adami DMV, MRCVS, DACVAA, DECVAA, RCVS and EBVS® European Specialist in Anaesthesia and Analgesia, PhD, Department of Clinical Sciences and Services, Royal Veterinary College, Hawkshead Lane, AL97TA, Hatfield, UK

Email: cadami@rvc.ac.uk

Abstract

Objectives To compare the Electronic von Frey Anaesthesiometer (EVF) and the Small Animal ALGOmeter (SMALGO), used to measure sensory thresholds in 13 healthy cats at both the stifle and the lumbosacral joint, in terms of inter-rater and inter-device reliability.

Methods Two independent observers carried out the sets of measurements in a randomised order, with a 45-minute interval between them, in each cat. The inter-rater and inter-device reliability were evaluated by calculating the inter-rater correlation coefficients (ICC) for each pair of measurements. The Bland-Altman method was used as an additional tool to assess the level of agreement between the two algometers.

Results The sensory thresholds measured with the EVF were 311 ± 116 g and 378 ± 178 g for the stifle and for the lumbosacral junction, respectively, whereas those measured with the SMALGO were 391 ± 172 g and 476 ± 172 g. The inter-rater reliability was fair (ICC > 0.4) for each pair of measurements except those taken at the level of the stifle with the SMALGO, for which the level of agreement between observer A and B was poor (ICC = 0.01). The inter-device reliability was good (ICC = 0.73; P= 0.001). The repetition of the measurements affected reliability, as the thresholds obtained after the 45 minute break were consistently lower than those measured during the first part of the trial (P = 0.02).

Conclusions and relevance The EVF and the SMALGO may be used interchangeably in cats, especially when the area to be tested is the lumbosacral joint. However, when the thresholds are measured at the stifle, the inter-observer reliability is better with the EVF

than with the SMALGO. The reliability decreases when the measurements are repeated within a short time interval, suggesting a limited clinical applicability of quantitative sensory testing with both algometers in cats.

Introduction

Recognizing and treating pain in feline patients has always been extraordinarily challenging. Traditionally, behavioural indicators are used to evaluate pain in cats,[1,2] and various species-specific pain scales have been developed on the basis of such indicators with the purpose of ameliorating peri-operative pain management.[3,4] Recently, the use of facial expressions as an additional tool to assess acute pain has become popular also in feline patients.[5]

Whilst for the evaluation of peri-operative acute pain the veterinarians can rely on a number of available and validated tools, scoring chronic pain remains a challenge even for the most experienced observers. Despite the lack of a unanimously accepted characterization of chronic pain in cats,[6] as a matter of fact cats do suffer from clinical conditions, such as osteoarthritis (OA),[7] which in humans and dogs are known to cause maladaptive pain.[8-10] In an attempt to evaluate OA-related feline pain, Benito and colleagues[11] developed and validated a Feline Musculoskeletal Pain Index (FMPI), based on subjective assessments performed by the owner in the animals' natural environment. With the same purpose, another study proposed the combined use of more objective parameters, namely gait analysis variables and mechanical sensory thresholds measured with an algometer.[12] Similarly, the Montreal Instrument for Cat Arthritis Testing, developed by Klinck and colleagues,[13] relies on a combination of behavioural indicators, mechanical thresholds and gait analysis.

The use of mechanical sensory thresholds as a tool to quantify chronic pain in cats is not novel, with most of the previous investigations that focused on this aspect relying on the use of the Electronic von Frey Anaesthesiometer (EVF).[13-15] This algometer is composed of a control unit and a sensory probe, used to apply over the body surface a force that is measured, displayed and stored. The force at which a predefined behavioural response is evoked is defined as threshold. Whilst the EVF has been designed for use in human patients, the Small Animal ALGOmeter (SMALGO), which shares with the former the working principle, has been specifically developed for laboratory rodents, and may represent a valid alternative to the EVF. The SMALGO was found useful and reliable to quantify pain in rats and mice in various experimental models, including inflammatory pain, mechanical allodynia and hyperalgesia.[16-18]

The primary aim of this study was to compare the EVF and the SMALGO, used to measure mechanical sensory thresholds in a population of healthy cats, at two anatomical sites commonly affected by feline OA, in terms of inter-rater and inter-device reliability. Secondary aims were to determine the effect of the repetition of a whole set of measurements, after a 45 minute-interval, on the reliability of both algometers, and to determine baseline mechanical sensory thresholds in healthy cats.

The authors hypothesized that the EVF and the SMALGO would be comparable for the use intended in this study, and that both inter-rater and inter-device reliability would be fair.

Materials and methods

Ethical approval

The study was conducted under ethical approval of both the University of Turin (Protocol number: 1245/120618) and the Clinical Research Ethical Review Board of the Royal Veterinary College of the University of London (License number: URN 2018 1773-3). A signed informed owner consent was obtained for each cat.

Animals

Thirteen cats, owned by either veterinarians or students in their 5th year of veterinary medicine, were enrolled in this study. The sample size was determined with the method described by Walter and colleagues (1998) for reliability studies, with the variables set as follows: number of observers = 2; desired value for inter-class correlation coefficient (ICC) = 0.8; minimally acceptable value for ICC = 0.05; $\alpha = 0.05$ and $\beta = 0.2$. This resulted in a minimal number of observations (cats) equal to 10. The Exclusion criteria were history of orthopaedic and neurological conditions that may have altered the sensory thresholds, and medical therapy with any drug with known analgesic effect. The cats were admitted to the Veterinary teaching Hospital of the University of Turin on the morning of the data collection, and left undisturbed for acclimatization in the examination room where the measurements were carried out for at least 15 minutes. Demographic data collected and used for statistics were sex, breed, age (months), body Condition Score

(BCS: 0-9),[20] body weight (kg) and height (cm), the latter measured from the dorsal end of the scapular spine, identified by palpation, to the surface of the examination table, with the cat in standing position. Food and water remained available until the trial was commenced.

Preparation of the instruments

Both devices are calibrated at the factory and do not require recalibration prior to use. However, before each set of measurements, the EVF was checked for accuracy as follows. After the 1000 g probe was equipped with a new rigid tip, a standard 5.3 g weight provided by the manufacturer was applied onto the tip, with the unit in horizontal position. The measurements were allowed to begin only in case the reading displayed and stored by the unit was equal to 5.3 ± 0.1 g. Regarding the SMALGO, the probe was equipped with the 3 mm sensor tip and the unit selected (g); following, the control unit was zeroed by resetting the tar to zero with a foot switch, and the key “max” pressed to allow the device to store the maximum force value recorded during the measurement.

Sensory thresholds measurements

Two anatomical sites were investigated: the lumbosacral intervertebral joint and the medial site of the stifle. The former was identified by using as anatomical landmarks the ileum wings, the last lumbar vertebra and the sacrum. For the latter, the target was the medial aspect of the knee, between the patella (dorsal) and the tibial tuberosity (ventral).

For both sites, the sensory tips of both instruments were applied perpendicularly to the skin, and a steadily increasing force applied until a positive behavioural reaction could be evoked. Attempt to escape, tail wiggling, hissing, attempts to bite or aggressions, ears back and flat against the head, head turning towards the site of stimulation, back muscle contraction (for the lumbosacral) and limb withdrawal (for the stifle) were considered positive behavioural reactions. When at least one of these reactions was observed, the mechanical stimulation was interrupted and the sensory tip withheld; the maximal force value displayed by the control unit was manually recorded. Each single measurement was repeated once to confirm the threshold, with a time interval of at least 30 seconds in order to avoid temporal summation;[21] the average calculated from these values was used for statistical analysis. Two observers (EL, observer A and CA, observer B) carried out the measurements independently, with the cats minimally restrained by the owner. A 45 minute-time interval was allowed between the subsequent sets of measurements carried out by the two observers. For each cat, the order of the observers and, for each observer, of the device to be used first and of the anatomical site to be assessed first, was determined by simple randomization based on flipping of a coin.

Statistics

Data distribution was assessed with both the Kolmogorov-Smirnov test and the Shapiro-Wilk test. The Spearman correlation coefficient (SCC) was calculated to detect correlations between the sensory thresholds and demographic variables of the cats (age, BCS, body weight and height). The inter-observer reliability was evaluated between observers A and B and, for each cat, between the observer who started the trial (1st observer) and the other, who carried out the measurements after the 45 minute break (2nd observer). The levels of agreement were quantified by calculating the inter-rater correlation coefficient (ICC), with 95% confidence intervals (CI; upper and lower bounds). The inter-device reliability was evaluated with both the ICC (with CI) and the Bland-Altman analysis.[22] A paired-T test was run to compare sets of measurements showings means and standard deviations that appeared to be different at first sight (between observer A and B, and between the 1st and the 2nd observer). P values < 0.05 were considered statistically significant. The level of agreement (both inter-observer and inter-device) was scored as follows: ICC < 0.40= poor; ICC between 0.40 and 0.59= fair; ICC between 0.60 and 0,74= good; and ICC between 0.75 and 1= excellent.[23] Commercially available softwares were used (IBM SPSS Statistics 24, IBM Corporation, NY, USA; and SigmaPlot 14 and SigmaStat 4, SYSTAT Software Inc, CA, USA).

Results

Normally distributed data are here presented as means and standard deviations, while data with non-normal distribution as medians and maximum-minimum ranges.

Twelve cats completed the study. One cat appeared to be stressed after the first set of measurements with the SMALGO, therefore it was decided to let him rest for about a hour and then allow the second observer to proceed with the measurements only with the SMALGO, in order to use these two sets of data for comparison.

Five cats were spayed females while the remaining 8 cats were neutered males. The represented breeds were domestic short hair (n=12) and domestic long hair (n=1). The cats were aged 60 [12-180] months, weighed 5.4 ± 1.2 kg, had a BCS of 5 [4-9] and their height was 28 ± 3.6 cm. There were significant positive correlations between both the body weight and the BCS, and the sensory thresholds (SCC: 0.21 and 0.27; and $P = 0.04$ and 0.007, respectively), and significant negative correlation between the height of the cats and the sensory thresholds (SCC: -0.31; $P = 0.001$). No correlation was found between the age of the cats and their sensory thresholds.

Observer A carried out the first set of measurements in 8 cats, while observer B started the trial in the remaining 5. There were no statistically significant differences between the sensory thresholds recorded by observers A and B, with both devices and at both anatomical sites. Overall, the thresholds recorded during the first set of measurements by one of the two observers (1st observer) with both devices and at both sites were significantly higher than those carried out by the other observer after the 45 minute break (2nd observer) (Table 1; $P = 0.02$). The level of agreement between these sets of measurements was poor (Table 2). The overall inter-rater agreement between 1st and 2nd observer was poor; however, when investigated in details, such agreement was fair when the measurements were carried out with the EVF at both the anatomical sites, and with the SMALGO at the lumbosacral joint, but poor for the measurements obtained with the SMALGO at the stifle (Table 2). The inter-device reliability was good (Table 2; P value: 0.001), although the level of agreement between the EVF and the SMALGO was better at the lumbosacral junction compared to the stifle, as demonstrated by the higher ICC obtained at the former site (Table 2). Data for each variable are presented in Table 1; the ICC for each set of comparison, together with the corresponding 95% CI, are shown in Table 2.

Figure legend

Figure 1 The Bland-Altman plot shows the difference between the thresholds measured with the EVF and those with the SMALGO (g) in 13 healthy cats, plotted against the average of all the measured thresholds.

Discussion

This study demonstrates that the measurement of sensory thresholds in healthy cats with both the SMALGO and the EVF does not result in consistent readings when the measurements are repeated after a relatively short time interval. In each cat, the repetition of the trial 45 minutes after the first set of measurements resulted in decreased sensory thresholds, which seems to indicate that the cats easily became sensitized or less cooperative after manipulation. Since a useful method to quantify pain should be repeatable in order to evaluate the efficacy of the analgesic therapy and titrate it to effect, this drawback limits the clinical applicability of quantitative sensory thresholds in feline patients. It also suggests that, if repeated tests are to be performed, a time interval longer than 45 minutes between subsequent measurements may help to improve reliability.

The good inter-device reliability indicates that the thresholds measured with the two algometers are similar, and suggests that both the EVF and the SMALGO might be used interchangeably in cats. However, comparable results are more likely to be obtained when the two algometers are used to measure sensory thresholds at the lumbosacral junction than at the level of the stifle. Moreover, both observers obtained higher thresholds with the SMALGO compared to the EVF. A possible explanation for this finding could be that the 3 mm sensory tip, chosen by the authors for the SMALGO, is too small for cats and needs a greater application force than the EVF probe to evoke comparable behavioural reactions. The 3 mm tip was chosen over the 5 and 8 mm ones as our clinical experience suggested that the former, owing to the pointed tip that applies the force on a small surface area, would evoke more consistent reactions than the flat 5 and 8 mm tips in cats.

Although the overall inter-rater agreement was poor, when this variable was analysed in details it showed that the agreement between observer A and observer B was fair for all pairs of measurements except the ones taken at the stifle with the SMALGO. The very poor agreement of this single comparison significantly affected the overall inter-rater agreement calculated between observer A and observer B, and could have been caused by a number of factors, including inappropriate selection of the SMALGO sensory tip, of the anatomical site, or both.

Investigating the feasibility of sensory thresholds as possible clinical tool to quantify, in the next future, pain in cats with degenerative joint disease was one of the focuses of this study. As a result, the stifle and the lumbosacral joint were chosen by the authors as anatomical sites of interest owing to their common involvement in feline osteoarthritis.[15,24,25] However, both investigators found the feline stifle a challenging anatomical site in terms of approachability when the cats were standing, and consistency and repeatability of the positioning of the sensory tip and subsequent application of the force. Regarding the future use of the EVF and of the SMALGO in the clinical setting, it is worth to consider that one of the intrinsic limitations of the current study is that its findings do not allow any conclusive statement about the validity of both devices for measuring pain in cats with actual OA.

Interestingly, physical variables of the cats, such as the height, the body weight and the BCS, had an effect on the sensory thresholds, which were higher in fat and heavier cats, and lower in taller, larger cats. Whilst the former finding could be due to the dampening effect of the adipose tissue covering both the lumbosacral joint and the stifle, which could have increase the tolerance of the cats to the mechanical stimulation in the area, providing a reasonable explanation for the inverse relationship between height and sensory threshold is more challenging. It might be hypothesized that large sized cats are more prone to develop osteoarthritis owing to the increased load on the joints, and that some of the taller cats of this study were affected. One study found that large breed cats, such as Maine coon, are prone to develop hip dysplasia.[26] However, whilst obesity and elderly are recognized risk factor for feline OA,[27] there is no published evidence that the size of the cats may act as well as predisposing condition. On the other hand, in this current study fatter cats had higher sensory thresholds, which indicates a higher tolerance to mechanical stimulation, and no correlation was found between sensory thresholds and elderly. The cats of the current study were owned by either a veterinarian or a veterinary medicine student, and regularly underwent clinical exam on occasion of standard vaccinations and deworming. Moreover, all owners were caring to their cats and it is reasonable to assume that they would notice changes in behaviour or signs of severe pain. Nevertheless, owing to the lack of a thorough orthopaedic and radiographic examination, the presence of osteoarthritis cannot be ruled out.

Conclusions

The good inter-device reliability suggests that the EVF and the SMALGO may be used interchangeably in cats; nevertheless, the poor inter-rater reliability observed when the SMALGO was used at the stifle indicates that, for this anatomical site, the EVF may represent a better option. Repetition of the measurements within a short time interval does affect reliability, a drawback that may limit the applicability of quantitative sensory testing with both algometers in clinical feline patients.

Acknowledgements The authors would like to thank Dr. Loris Barale, the University of Turin and all the cats' owners for their help with this study.

Conflict of interest The authors declared no potential conflict of interest with respect to the research, authorship, and/or publication of this article.

References

1. Merola, I, Mills, DS. **Behavioural signs of pain in cats: an expert consensus.** *PLoS One* 2016; 11: DOI: 10.1371/journal.pone.0150040.
2. Merola, I, Mills, DS. **Systematic review of the behavioural assessment of pain in cats.** *J Feline Med Surg* 2016; 18: 60–76.
3. Reid J, Scott EM, Calvo G et al. **Definitive Glasgow acute pain scale for cats: validation and intervention level.** *Vet Rec* 2017; 180: 449.
4. Brondani, JT, Mama, KR, Luna, SP. **Validation of the English version of the UNESP-Botucatu multidimensional composite pain scale for assessing postoperative pain in cats.** *BMC Vet Res* 2013; 9: 143.
5. Holden E, Calvo G, Collins M et al. **Evaluation of facial expression in acute pain in cats.** *J Small Anim Pract* 2014; 55: 615-621.
6. Grubb T. **What do we really know about the drugs we use to treat chronic pain?** *Top in Companion Anim Med* 2010; 25: 10-19.
7. Lascelles DB, Dong YH, Marcellin-Little DJ, et al. **Relationship of orthopedic examination, goniometric measurements, and radiographic signs of degenerative joint disease in cats.** BMC Veterinary Research 2012; 8:10.
8. Dimitroulas T, Duarte RV, Behura A et al. **Neuropathic pain in osteoarthritis: a review of pathophysiological mechanisms and implications for treatment.** *Semin Arthritis Rheum* 2014; 44: 145-154.

9. Gagnon A, Brown D, Moreau M et al. **Therapeutic response analysis in dogs with naturally occurring osteoarthritis.** *Vet Anaesth Analg* 2017; 44: 1373-1381.

10. Knazovicky D, Helgeson ES, Case B et al. **Replicate effects and test-retest reliability of quantitative sensory threshold testing in dogs with and without chronic pain.** *Vet Anaesth Analg* 2017; 44: 615-624.

11. Benito J, Depuy V, Hardie E, et al. **Reliability and discriminatory testing of a client-based metrology instrument, feline musculoskeletal pain index (FMPI) for the evaluation of degenerative joint disease-associated pain in cats.** *Vet J* 2013; 196: 368-373.

12. Guillot M, Moreau M, Heit M et al. **Characterization of osteoarthritis in cats and meloxicam efficacy using objective chronic pain evaluation tools.** *Vet J* 2013; 196: 360-367.

13. Klinck MP, Rialland P, Guillot M et al. **Preliminary validation and reliability testing of the Montreal Instrument for cat arthritis testing, for use by veterinarians, in a colony of laboratory cats.** *Animals (Basel)* 2015; 5: 1252-1267.

14. Addison ES and Clements DN. **Repeatability of quantitative sensory testing in healthy cats in a clinical setting with comparison to cats with osteoarthritis.** *J Feline Med Surg* 2017; 19: 1274-1282.

15. Stadig S, Lascelles BD, Bergh A. **Do cats with a cranial cruciate ligament injury and osteoarthritis demonstrate a different gait pattern and behaviour compared to sound cats?** *Acta Vet Scand* 2016; 58:71-79.

16. Kim JS, Ahmadinia K, Li X et al. **Development of an Experimental Animal Model for Lower Back Pain by Percutaneous Injury-Induced Lumbar Facet Joint Osteoarthritis.** *J Cell Physiol* 2015; 230: 2837-2847.

17. Girard P, Verniers D and Coppé MC. **Nefopam and ketoprofen synergy in rodent models of antinociception.** *Eur J Pharmacol* 2008; 584: 263-271.

18. Reynoso-Moreno I, Najar-Guerrero I, Escareño N, et al. **An endocannabinoid uptake inhibitor from black pepper exerts pronounced anti-inflammatory effects in mice.** *J Agric Food Chem* 2017; 65: 9435-9442.

19. Walter SD, Eliasziw M, Donner A. **Sample size and optimal designs for reliability studies.** *Stat Med* 1998; 17:101-110.

20. LaFlamme DP. **Development and validation of a body condition score system for cats: a clinical tool.** *Feline Pract* 1997; 25: 13-18.

21. Nie H, Arendt-Nielsen L and Andersen H. Temporal summation of pain evoked by mechanical stimulation in deep and superficial tissue. J Pain 2005; 6: 348-355.

22. Bland JM and Altman DG. **Statistical methods for assessing agreement between two methods of clinical measurement**. *Lancet* 1986; 327: 307–310.

23. Cicchetti DV. **Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology**. *Psychological Assessment* 1994; 6: 284–290.

24. Clarke SP, Mellor D, Clements DN, et al. **Prevalence of radiographic signs of degenerative joint disease in a hospital population of cats.** *Vet Rec* 2005; 157: 793-799.

25. Lascelles BD, Henry JB, Brown J, et al. **Cross-sectional study of the prevalence of radiographic degenerative joint disease in domesticated cats.** *Vet Surg* 2010; 39: 535-544.

26. Keller GG, Reed AL, Lattimer JC, et al. **Hip dysplasia: a feline population study.** *Vet Radiol Ultrasound* 1999; 40: 460-440.

27. Hardie EM, Roe SC, Martin FR. **Radiographic evidence of degenerative joint disease in geriatric cats: 100 cases (1994-1997)**. *J Am Vet Med Assoc* 2002; 220: 628-632.