TITLE: Development and initial validation of a sensory threshold examination protocol (STEP) for phenotyping canine pain syndromes

AUTHORS: Sandra Sanchis-Mora, Yu-Mei Chang, Siobhan Abeyesinghe, Amy Fisher, Holger A. Volk, Ludovic Pelligand

JOURNAL: Veterinary Anaesthesia and Analgesia

PUBLISHER: Elsevier

PUBLICATION DATE: 24 February 2017 (online)

DOI: 10.1016/j.vaa.2016.09.004

RESEARCH PAPER

Development and initial validation of a sensory threshold examination protocol (STEP) for phenotyping canine pain syndromes

Sandra Sanchis-Mora[a,*], Yu-Mei Chang[b], Siobhan Abeyesinghe[c], Amy Fisher[d], Holger A Volk[a] & Ludovic Pelligand[a, e]

[a]Comparative Biomedical Sciences, Royal Veterinary College, University of London, UK.

[b]Research Office, Royal Veterinary College, University of London, UK,

[c]Production and Population Health, Royal Veterinary College, University of London, UK

[d]Transpharmation Ltd, The London Bioscience Innovation Centre, 2 Royal College Street, London, UK

[e]Clinical Science and Services, Royal Veterinary College, University of London, UK.

[*]**Correspondence:** Sandra Sanchis-Mora, Royal Veterinary College, Hawkshead Lane, North Mymms, Herts, AL9 7TA, E-mail: ssanchismora@rvc.ac.uk

Running head: Quantitative sensory testing protocol in healthy dogs

Abstract

Objective To study feasibility and test-retest repeatability of a sensory threshold examination protocol (STEP) and report quantitative sensory threshold distributions in healthy dogs.

Study design Prospective, observational, cohort study.

Animals Twenty-five healthy client-owned dogs.

Methods Tactile sensitivity (TST) (von Frey filaments), mechanical thresholds (MT with 2, 4 and 8 mm probes), heat thresholds (HT) and responsiveness to cold stimulus (CT at 0℃) were quantitatively assessed for five body areas (BA: tibias, humeri, neck, thoracolumbar region and abdomen) in a randomized order on three different occasions. Linear Mixed Model and Generalised Linear Mixed models were used to evaluate the effects of body weight category, age, sex, BA, occasion, feasibility score and investigator experience. Test-retest repeatability was evaluated with the Intra-class Correlation Coefficient (ICC).

Results The STEP lasted 90 minutes without side effects. The BA affected most tests ($p \leq 0.001$). Higher thresholds and longer cold latencies were scored in the neck ($p \leq 0.024$) compared to other BAs. Weight category affected all thresholds ($p \leq 0.037$). Small dogs had lower MT (~1.4 N mean difference) and HT (1.1 ^{0}C mean difference) than other dogs ($p \leq 0.029$). Young dogs had higher HT than adults (2.2 ^{0}C mean difference) ($p = 0.035$). Gender also affected TST, MT and HT ($p < 0.05$) (females *versus* males: TST OR= 0.5, MT= 1.3 N mean difference, HT= 2.2 ^{0}C mean difference). Repeatability was substantial to moderate for all tests, but poor for TST. There was no difference in thresholds between occasions, except for CT. Test-retest repeatability was slightly better with the 2 mm MT probe compared to other diameters and improved with operator experience.

Conclusions and clinical relevance The STEP was feasible, well tolerated and showed substantial test-retest repeatability in healthy dogs. Further validation is needed in dogs suffering pain.

Keywords dog, mechanical, nociception, quantitative sensory testing, thermal

Introduction

Quantitative sensory testing (QST) is a method used to quantify the somatosensory function (Backonja et al. 2013; Edwards et al. 2016). In the clinical research setting, QST evaluation comprises touch and vibration detection, as well as mechanical and thermal (heat and cold) noxious stimuli (Walk et al. 2009). In humans, QST has been applied in healthy volunteers, patients with neuropathic pain defined as "pain caused by a lesion or disease of the somatosensory nervous system" (Backonja et al. 2013; Finnerup et al. 2016) and other pain syndromes where the somatosensory function may be altered due to peripherial or central sensitisation (Whitaker et al. 2016).

There is evidence in people that altered somatosensory function originates from various pathophysiological mechanisms that can be elucidated by the results of a QST panel (Greenspan 2001; Hansson 2002; Hansson et al. 2007). The QST may identify patient subgroups with certain underlying neuro-biological mechanisms who may respond differently to a given drug (Baron et al. 2014). Characterizing the somatosensory phenotype of patients with chronic pain by identifying sensory abnormalities (positive, such as hyperalgesia and allodynia or negative, such as numbness or lack of sensation), is necessary to help select the best therapeutic class for a specific patient. This is the key to mechanism-based diagnosis and could significantly improve treatment (Rolke et al. 2006; Reimer et al. 2014; Edwards et al. 2016).

Similar to humans, animals experience chronic pain of neuropathic origin (Mathews 2008). The QST has the potential to be a neurophysiological tool in veterinary medicine and has been used in different clinical and experimental models such as osteoarthritis, hip replacement and ovariohysterectomy in dogs (Brydges et al. 2012; Hunt et al. 2013; Moore et al. 2013; Tomas et al. 2014). Recently, thermal stimuli have been tested in combination with

mechanical stimuli in canine models of osteoarthritis and spinal cord injury (Knazovicki et al. 2016; Gorney et al. 2016; Song et al. 2016). However, the combination of all the QST modalities together in one standardized test has never been explored.

The use of naturally occurring canine pain models is becoming a valuable option to study human chronic pain (Lascelles 2013). They better mirror human conditions and may provide better insight into drug efficacy in humans compared with experimentally induced rodent models. Observing the responses of dogs administered analgesic drugs for different naturally occurring pathophysiologic mechanisms are powerful models for translational studies. Designing a standardized method to evaluate nociceptive thresholds in canine patients and defining sources of confounding factors in healthy dogs will ultimately offer an improvement of diagnosis and characterisation of chronic pain.

The aims of this study were to (i) evaluate the feasibility and test-retest repeatability of a QST sensory threshold examination protocol (STEP) including tactile, thermal and mechanical testing; (ii) to identify explanatory variables affecting results; and (iii) to provide baseline QST thresholds and their distribution in a sample of healthy dogs for its use as a tool to phenotype chronic pain syndromes in future studies.

Materials and Methods

The project was approved by the Royal Veterinary College Ethics and Welfare Committee (URN 2013 1243). Twenty-five healthy client-owned dogs were included in the study which was conducted between January and August 2014. Signed owner consent was obtained for all animals enrolled in the study. The dogs were deemed healthy based on their medical history and a complete physical/neurological/orthopaedic exam performed by a veterinarian. Owners completed the Canine Brief Pain Inventory (CBPI) which consisted of three parts: pain severity ranging from 0 (no pain) to 10 (extreme pain), pain interference from 0 (no interference) to 10 (completely interferes) and quality of life assessment from 1 (poor) to 5 (excellent) (Brown et al. 2008). An inclusion criterion was a CBPI score of 0 on pain severity and pain interference, with a quality of life scores greater than 4 (very good) (Brown et al. 2008). Dogs that were not able to attended a minimum of two appointments (occasions) were excluded.

Animals were tested on 2 or 3 occasions (occasion 1, 2 or 3) (Fig. 1), each separated by a week, with a sensory threshold examination protocol (STEP). The CBPI was completed on each occasion to ensure that no changes occurred over time in order to continue remain in the study. The standardised STEP consisted of a tactile sensitivity test (TST using von Frey filaments), mechanical thresholds (MT using a calibrated veterinary pressure algometer), heat and cold thresholds (HT, CT). The tests were applied in the same order in all dogs as follows: TST, MT, HT, CT. Mechanical testing was performed before thermal to avoid iatrogenic sensitisation, according Grone et al. (2012). For each sensory modality, measurements were taken from five different body areas (BA) in a randomized order (www.graphpad.com/quickcalcs): bilaterally over the mid tibias, mid humeri, neck area, thoraco-lumbar (T-L) area and left side only over the abdomen (Fig. 2). Dogs were all tested in the same room in standing position. Prior to testing, dogs were acclimatised to the room for

five minutes before clipping. Clipping of the BAs (1.5 x 1.5 cm patch) was needed to allow TST and thermal evaluation. The areas were clipped on each occasion. The test started not less than ten minutes after clipping.

Each individual test terminated with the observation of one of the following endpoints: turning the head towards the device, growling, lip licking, or backing away from the stimulus. A feasibility score ranging from 1 (no problem) to 5 (impossible) adapted from Briley et al. (2014) (Appendix 1) was used to evaluate dog cooperation. All of the tests were readily escapable and, if an animal appeared to be in discomfort during testing (or unable to tolerate the protocol), the test was terminated immediately. If the dogs showed fatigue or reluctance to stand, time was allowed for resting of up to five minutes between tests.

Tactile sensitivity thresholds

Von Frey filaments (20 filaments, 0.008 to 300 gram force (gf); Bioseb, France) were used for TST. The hairs were pressed against the skin with enough force so that the hair buckled and formed a U-shape. Two techniques were applied and compared. First, a group of 18 dogs were tested with the *up-down technique* (TST_{UD}) described by Chaplan et al. (1994). The test was initiated with an intermediate 2.0 gf hair. A lack of response to a filament dictated that the next thickest filament was used in the following stimulation ('up rule'), while a positive response dictated the use of the next thinnest filament ('down rule'). When the animal first changed its response pattern: a negative response followed by a positive response or vice versa, another four von Frey presentations were done according to the above 'up-down rules'. The final response threshold was interpolated using the formula: gf threshold = (10 [Xf+kδ])/10,000 where Xf = value (in log units) of the final von Frey filament used; k = tabular value (see Chaplan et al. 1994 for more details) for the pattern of positive/negative responses; and δ = mean difference (in log units) between stimuli.

The *50% response technique* ($TST_{50\%}$) described by Brydges et al. (2012) was used in a second group of 7 dogs, because preliminary data from the *up-down technique* suggested difficulties in interpretation of the final threshold as a result of data censoring (animals not responding to the thickest filament). The $TST_{50\%}$ consisted of using the filaments in ascending order. Each filament was applied six times, with 3 second intervals. If no aversive response was obtained after testing with a small diameter filament, the next highest diameter filament was used. The tactile sensitivity threshold was defined by the filament that first induced a withdrawal response at least three times in six repeated measurements.

Mechanical thresholds

Mechanical response was tested with a calibrated veterinary pressure algometer (ProdPro; Topcat Metrology Ltd, UK), equipped with three different probe diameters: 2 mm, 4 mm, and 8 mm. The accuracy of the instrument was ± 0.5 Newton (N) within a range of 0.5−25 N. The algometer provided a constant increment pressure increase of 2 N $second^{-1}$ to achieve repeatable applications. The device was applied perpendicular to the skin of the dogs with one hand. The other hand was used to support gently the medial aspect or the contralateral side of the area tested. Three repetitions in the five BAs were obtained for each occasion with the three different probe sizes. Twenty seconds were allowed between repetitions. The final thresholds for the occasion were obtained calculating the mean of the three repeats per BA.

Thermal thresholds

Heat stimulus was applied using a veterinary thermal probe (HotPro; Topcat Metrology Ltd). The device was a handheld calibrated prototype adapted from the already validated wired version (Dixon et al. 2002). Before testing, the skin temperature was measured with the device and room temperature was recorded (EL-USB-TP-LCD; Lascar Electronics, UK). During testing, the temperature increased from baseline to a maximum of 55 ^{0}C with a ramp

of 1 ^{0}C second^{-1} until the endpoint was reached. The device was applied as described in the use of the pressure algometer. Three repetitions in the five BAs were obtained for each occasion. Twenty seconds were allowed between repetitions. The final threshold for the same occasion was obtained calculating the mean of the three repeats per BA.

Cold stimulus was applied using a handheld thermal probe (NTE-2A; Physitemp Instruments, NJ, USA) with a 13 mm diameter surface set at 0 ± 0.2 ^{0}C. The probe used a peltier semiconductor heat pump and a digital temperature control unit to maintain accurate temperature application during trials. The latency (seconds) between application and observation of endpoint was recorded. Three repetitions for each BA were obtained on each occasion. Each repeat included the entire series of BAs in a randomized order, starting again the entire series in the same random order for the second and the third repeat. This allowed at least 60 seconds between repeats in the same BA maintaining appropriate duration of the total time spent in all the tests.

Analysis of data

Data were analysed using statistical software (IBM SPSS 21). Data from dogs which the feasibility scores were higher than 2 were excluded from the analysis. For continuous data, normality of distribution was verified by Kolmorov-Smirnov's test and by visual assessment of Q-Q plots and histograms. When required, data were logarithmically transformed to verify the assumption of data normality prior to parametric testing. Cold and tactile sensitivity thresholds were right-censored (60 seconds and highest filament, respectively) and treated as binary data (0 = response below threshold and 1 = threshold reached).

Continuous data were expressed as mean ± standard deviation (SD). Data following a logarithmic distribution were presented as geometric mean and back-transformed SD. Other data were presented as median (range). For graphical display, median, interquartile range and

minimum-maximum was used. Categorical data were expressed as number out of total and percentage. Significant differences were considered if $p < 0.05$.

Data were divided in two periods of testing (first period of testing form January 2014 to April 2014 against second period of testing from May 2014 to August 2014) to evaluate the effect of the operator gaining experience with QST thresholds.

A linear mixed model (LMM) was used for continuous outcome variables MT (N) and HT (°C) separately, to evaluate the influence of the explanatory variables on the within/between subject variability. Subjects were considered as a random effect. The following explanatory variables were considered as fixed effects: body weight and age (divided in three categories respectively, Appendix 2), sex, BA (5 total), right/left side. Analysis of HT also included body temperature and room temperature as additional fixed effects. Factors affecting the metrological performance of the protocol were also included in the model as fixed effects: feasibility score (0, 1 or 2), effect of repeated testing (occasion 1, 2 or 3) and period of testing (first and second period). In the case of the pressure algometer, the three different probes (2, 4 and 8 mm) were compared in separate statistical models (MT2, MT4, MT8). Magnitude of the effects was reported as the adjusted mean difference and p-value.

A generalised Linear Mixed Model (GLMM) was used for tests with binary logistic outcomes (TST and CT). The dependent variables were response to any of the von Frey filaments and 0 ^{0}C before 60 seconds (pTST and pCT) respectively. The fixed and random effects were the same as for continuous outcomes. Magnitude of the effects was reported as the odds ratio and *p* value.

Interactions were evaluated when appropriate. Post-hoc comparisons of the significant effects were made using Fisher's least standard differences (LSD) method.

Test-retest repeatability was evaluated by calculating the intra-class correlation coefficient (ICC). The ICC is the degree of closeness of repeated measures in a group of individuals (Andersen et al. 2014). It describes the contribution of the variation within the individual within the total variation (between dogs variation + within dogs variation + error variation) (Vangeneugden et al. 2004). Therefore, the closest to 1 the ICC, the smallest the variation within dogs across the different occasions (occasion 1, 2 or 3), and the better the repeatability of the test. The ICCs were categorised as slight/poor (< 0.2), fair (>0 .2 to 0.4), moderate (> 0.4 to 0.6), substantial (> 0.6 to 0.8) and almost perfect (> 0.8) (Landis & Koch 1977).

Results

Descriptive results

The twenty-five healthy client-owned dogs included in the study (Fig.1) had an age of 6.0 (0.3−9.0) years and body weight of 15 (6−35) kg. There were 14 females (56%) and 11 males (44%). All dogs' CBPI scores were 0 for pain intensity and pain interference, and 5 for quality of life. Eleven dogs (44%) were tested during the first period of testing. Distributions of the sample by different weight category and age are shown in Appendix 2. Feasibility score distribution across the sample of dogs was 0 for 4 dogs, (16%); 1 for 9 dogs, (36%); 2 for 12 dogs, (48%). The temperature of the testing room was 22.9 (19.3−26.2) ^{0}C. The skin temperature was 30.9 (27.6 − 33.2) ^{0}C. According to this range of skin temperature, the baseline starting temperature was set at 30^{0}C for HT in all dogs. The STEP protocol took 90 minutes per dog and was applied with no side effects reported by owners.

Mean ± SD or median (range) of the TST, MT, HT, and CT are displayed in Tables 1 and 2, respectively. Median (interquartile range) and minimum-maximum thresholds for the different stimuli are summarised for the different BA in Fig. 3.

Influence of explanatory variables

The *p* values of the different explanatory variables studied are summarised in Table 2. The post-hoc comparisons for these effects are reported in Appendix 3 (mean differences and *p* - value for MT and HT; odds ratio and p value for TST and CT). There was a highly significant effect of the BA tested for all stimuli evaluated ($p \leq 0.001$). The QST thresholds for the different BA and stimuli are summarised in Fig. 3. Higher thresholds were scored in the neck compared with other areas in all the QST ($p \leq 0.024$) (Appendix 3). Left and right sides of each BA showed no significant differences in thresholds in this study (Table 2).

Weight category had a significant effect on all thresholds ($p \leq 0.037$) except for p TST_{UD}. Small dogs had lower MT and HT than medium and large dogs ($p \leq 0.029$, Table S3). Nevertheless smaller dogs were less likely to respond to $TST_{50\%}$ than larger dogs ($p < 0.01$). Regarding age, young dogs were more likely to obtain higher HT than adults ($p = 0.035$), however, adults obtained lower HT than geriatric patients did ($p = 0.013$). The MT and HT were significantly higher in females ($p < 0.05$) whereas this effect was not significant for pCT. In contrast, $pTST_{50\%}$ was higher in females than in males ($p = 0.006$ and $p = 0.009$ for TST_{UD} and $TST_{50\%}$ respectively).

Test-retest repeatability

There was no inter-occasion difference, except for pCT (Table 2), where percentage of response was significantly higher during the last occasion than the previous two ($p < 0.01$). Feasibility score only significantly affected $pTST_{UD}$ ($p = 0.004$); higher proportion of responses was obtained with higher feasibility scores (less cooperative dogs). Lower thresholds were obtained for MT on the second period of testing where the operator obtained more experience ($p < 0.05$) (Appendix 3).

The ICCs showed moderate to substantial test-retest repeatability across occasions (Table 3) except for the TST_{UD} where the ICC was poor. The two periods of testing showed significant effect on MT. Therefore, the ICCs of the two periods for MT were calculated. A slight improvement in ICCs was seen (Table 3).

Discussion

Canine spontaneous models of chronic pain need a standard procedure for characterisation. In addition, investigations of nociception in animals should represent the preliminary step before clinical studies are undertaken to pursue better treatment options in small companion animals (Bergadano et al. 2006). This study intended to create and evaluate a sensory threshold examination protocol (STEP) to determine a complete QST phenotype in one clinical session. Feasibility, test-retest repeatability, and possible confounding factors (cofactors and covariates) to take into account when applying the STEP were studied.

First, consistently with other studies in dogs (Moore et al. 2013; Briley et al. 2014; Harris et al. 2015), the cofactor that had the largest effect in our study was weight category. Nevertheless, the sample in this study was not large enough to include weight as a continuous explanatory variable and the diversity of breeds was not representative enough to include this effect in the analysis. Another important factor affecting response is the limb length and the distance between the nociceptor to the brain (Blankenburg et al. 2010). Practically, thresholds obtained with the STEP should be compared between dogs of the same weight category.

Secondly, different BAs appeared to show very different thresholds, in line with other studies in healthy dogs (Coleman et al. 2014; Harris et al. 2015) and humans (Rolke et al. 2006). We included different body areas in this protocol so a map of QST thresholds could be evaluated for feasibility, test-retest repeatability and to evaluate if different body areas could show different thresholds as other studies have demonstrated. The choice of body areas in the present study was adapted from previous studies (Coleman et al. 2014; Harris et al. 2014) and modified to be performed easily with the tools provided) to ensure a good contact and avoid the probe slipping off the tested body area. This may allow different clinicians to use the STEP efficiently and with good results.

Neck area scored higher thresholds in all tests of the STEP. There are no other reports of neck thermal or mechanical testing in dogs. It has been suggested that tissues in the more distal aspects of limbs are more highly innervated than more proximal tissues and nerves have smaller receptive fields (Coleman et al. 2014). Contributing factors may also include differences in reaction time related to thickness of epidermis (Blankenburg et al. 2010). These findings support the assumption that when testing a patient for sensory abnormalities, thresholds from a specific BA should not be compared with values from a BA of a different location. The lack of differential sensitivity across the left and right sides suggests the unaffected side of a BA may be an appropriate control for the unilateral affected painful side if this has not been compromised by central sensitisation.

BAs significantly affected algometer readings in previous studies (Coleman et al. 2014; Harris et al 2014). Mechanical thresholds for spine and hips reported by Coleman et al (2014) (mean of approximately 38 N and 42 N, respectively) were higher than elbows and stifles (mean between 37 N and 27 N). It is difficult to compare these results to ours because the testing device differed and large dogs (retrievers) were tested in lateral recumbency; all of which could explain their high MTs (Coleman et al 2014). The same finding was reported in studies comparing healthy and osteoarthritic dogs in lateral recumbency (Knazovicky et al. 2016). The MTs on the tibia with a different device were higher when comparing within the same weight category range of our study (1523 gf being approximately 14.0 N *versus* 9.5 N obtained in our study with the 4 mm tip size). In this case, tip diameter was 3 mm and the rate of increase of pressure was not indicated. The MTs reported for the different body areas by Harris et al (2014) with the same device used in our study (i.e. MT of the tibias obtained a mean of 5.6- 5.8 N) were not separated by weight. Briley et al. (2014) obtained a mean between 1089 to 1028 gf, which corresponds with approximately 10 N. However, this was on

299 the metatarsal surface, in lateral recumbency and with a different algometer in healthy dogs
300 between 10 to 40 kg, which makes it impossible to compare between studies.

301 There are no other known veterinary studies reporting differences in BA in thermal
302 thresholds in dogs for direct comparison. Hoffman et al. (2012) reported a mean HT of 39 ^{0}C
303 on the lateral thorax in Beagles weighting 17 kg. Williams et al. (2014) measured the latency
304 of time healthy dogs were able to tolerate standing on a hot infrared light that reached about
305 59 ^{0}C in 30 seconds. Only the hind paw latency was evaluated in this study. Knazovicky et al.
306 (2016) applied a temperature of 45^{0}C on the tibias and other locations of the hind limb and
307 measured latency in large dogs. These areas were not clipped and prevent comparisons
308 between studies.

309 Previously, latency to respond to cold has been evaluated only on a cold plate at 6 ^{0}C in the
310 hind paw and the pelvic limb in lateral recumbency in healthy dogs (Brydges et al. 2012;
311 Briley et al. 2014) but not in thoracic limbs, neck or spine. Control dogs reached the cut off
312 time in most of the cases, as occurred in our study. Knazovicky et al. (2016) reported a mean
313 latency to 0^{0}C of 52.77 seconds in large dogs in lateral recumbency compared with a median
314 of 43.25 seconds obtained in the tibias in our study. Nevertheless, a standard methodology of
315 testing that allows good test-retest repeatability is necessary to establish a normal range and
316 allow comparison with chronic pain conditions in future studies.

317 Third, age affected the response to testing, as young and geriatric patients showed higher HTs
318 than adults did. Our results are consistent with human studies in which age differences had a
319 large effect in the data. (Rolke et al. 2006; Blankenburg et al. 2010). These effects could be
320 related with functional maturation of interneurons in the cortex and dorsal horn when
321 comparing young patients and decrease in innervation density when testing geriatric patients.

322 Fourth, the TST data in this study are in agreement with human studies showing that women

tend to be more sensitive to pain than men (Rolke et al. 2006). This has been also reported in dogs from the same breed when tested for MT (Coleman et al. 2014) and may be related with differences in central processing due to genetic and psychological factors (Blankenburg et al. 2010). However, our results showed the opposite pattern for MT and HT. This could be potentially explained, although not statistically significant, by the higher thresholds obtained by females in the younger group compare to adult group, especially on occasion 3.

The von Frey filaments determine a tactile sensory threshold, but not a nociceptive threshold. The TST assesses Aβ fibres (Hansson et al. 2007). For the TST, it was impossible to assess presence of mechanical allodynia since it was not present in the sample of healthy dogs tested and the % of response to any the von Frey filaments was very variable (Table 1). A similar pattern was observed with CT, where latency at which the cold stimulus (0 ^{0}C) may become nociceptive (assessment of Aδ and C fibres) could not be established due to the lack of response to cold in some dogs/BAs. The upper limits for HT and CT are actually the upper possible safety limits; therefore, a true upper range could not be obtained in this case (censored data). These problems have also been reported in healthy human volunteers (Rolke et al. 2006). Briley et al. (2014) studied the feasibility of the same device used in our study, demonstrating similarly to our finding large variability of response to 0 ^{0}C during the same cut off time, with healthy dogs. Dogs with osteoarthritis and spinal cord injury showed lower latencies to 0^{0}C compared to healthy dogs (Knazovicky et al. 2016; Gorney et al. 2016). However, further studies in dogs with different pain modalities are needed to elucidate whether this device could be used as a tool to detect allodynia or hyperalgesia, as it seems that 0 ^{0}C did not trigger a nociceptive response within 60 seconds in all healthy dogs.

Two methods to evaluate TST were compared in this study. The $TST_{50\%}$ has been used previously in dogs with cranial cruciate ligament rupture (Brydges et al. 2012) showing good results in identifying individuals with central sensitisation. These authors reported a mean of

348 900 mN mm^2 in control dogs between the second and the third digit of the hind limb, which
349 corresponds approximately with 300 gf; similar to our findings. It seemed that, although still
350 variable, a higher proportion of healthy dogs responded below the cut off with the $TST_{50\%}$.
351 The present study showed that the $TST_{50\%}$ technique was more repeatable, with less
352 variability between subjects and behaved similarly to other tests regarding factors influencing
353 results such as weight category, gender and body areas when compared with the TST_{UD}. In
354 contrast, the TST_{UD} did not have a good utility in healthy dogs.

355 For mechanical thresholds, methods of testing need standardisation as wider tip diameters
356 have been associated with higher thresholds and a large data range or between-individual
357 variability (higher SD) in previous studies (Harris et al. 2015; Taylor et al. 2015a). Our
358 results show similar ICCs for the different probe sizes with only slightly higher repeatability
359 using the 2 mm probe as previously reported (Harris et al. 2015). However, other studies used
360 different methods of assessment of test-retest repeatability (Harris et al. 2015; Taylor et al.
361 2015a).

362 In veterinary medicine, the reliability of QST has been assessed with different methods to
363 evaluate variation in QST thresholds over time (Williams et al. 2014; Brydges et al. 2012;
364 Moore et al. 2013; Briley et al. 2014; Gorney et al. 2016; Song et al. 2016). It has been
365 suggested that the most appropriate method to report test-retest repeatability when exploring
366 QST protocols (Moloney, 2012) is the ICC in conjunction with a measure of precision (i.e.
367 95% confidence interval). However, this method has its limitations, especially if the 95% CI
368 is large as occurred for TST and CT in this study. When the variability between individuals is
369 very large, it can also provide a falsely good ICC, and should be interpreted with caution (Lee
370 et al. 2012).

Chong and Cros (2004) defined QST evaluation as a subjective psychophysical test, where the consistency of the data relies on environmental factors, methodological factors and the attention and cooperation of the individual being tested. To help with this possible bias in our study, a feasibility score adapted from a previous study assessing mechanical and thermal thresholds in dogs in lateral recumbency (Briley et al. 2014) was used to evaluate cooperation of dogs and reaction to the stimuli. Feasibility score only affected $pTST_{UD}$, thus overall we found good cooperation > 50% of the time, mild sensitivity to being touched and mild variation in reaction to stimuli; sufficient to ensure a good feasibility and repeatability of the STEP. A higher proportion of dogs responded the 3rd testing occasion for CT, probably trying to avoid an uncomfortable sensation learned from previous tests. Other studies evaluating mechanical testing with other devices also showed a learning effect (Coleman et al. 2014).

An effect of the operator's experience was also evident for MT. During second period of testing, not only were MTs lower but also ICCs were slightly better compared with first period, and thorough operator training is advised before clinical use. Standardization of instructions to subjects, training of technicians, machine calibration, stimulus characteristics, and testing algorithms are all essential for accurate and reproducible QST (Chong & Cros 2004).

Protocols involving QST evaluation in humans include verbal communication of detection thresholds. In veterinary patients, this approach cannot be used and instead reliance must be placed on observable behavioural indicators. In the case of animals with peripheral and central sensitisation, where somatosensory function evaluated by QST encompasses the presence of allodynia or hyperalgesia as well as pain it is not possible to reliably distinguish between thresholds of sensation and nociception. Consequently, some authors view QST as a

395 semi-objective assessment (Gorney et al. 2016). Nevertheless, QST can provide valuable
396 clinical information regarding the impacts on patients (Brown 2012).

397 Limitations of the study include the small number of dogs tested. Further data may be
398 required to obtain reliable reference values. In future studies, dogs with inability to stand may
399 not be suitable for the current protocol. Position (sitting, laying in lateral recumbency) has
400 been tested in other studies (Harrys et al. 2014; Knazovivky et al. 2016; Gorney et al. 2016)
401 and could be a possibility for these patients. Fatigue from remaining standing was accounted
402 for and short periods of resting were allowed between tests. Clipping may not be possible in
403 some patients with severe allodynia, and the full battery of tests may not be possible to
404 perform in that particular body area: instead, other diagnostic tools could compliment the
405 assessment, including history, imaging tests, chronic pain questionnaires and behavioural
406 response when approaching the area.

407 In conclusion, the sensory testing examination protocol showed substantial to moderate test-
408 retest repeatability for HT and MT in healthy dogs. The STEP was feasible, safe and well
409 tolerated. Cold and tactile sensitivity thresholds showed poor consistency in response to the
410 stimuli and ICCs showed heterogeneity across these data. Further work in dogs with central
411 sensitisation is needed to assess the usefulness and test-retest repeatability of the STEP in
412 practice. Testing only the specific BA of interest could be envisaged to shorten the duration
413 of the protocol when phenotyping different pain conditions. Since weight category was the
414 most significant explanatory variable, nociceptive thresholds for the STEP were displayed
415 based on this covariate and in future should only be compared within weight class. Further
416 studies in dogs with painful conditions should evaluate the utility of each test in detecting
417 sensory abnormalities in dogs.

418 **Acknowledgements**

The authors would like to thank Heather Williams, Jo Murrell and Nicolas Granger from the University of Bristol for their collaborative support in this area of research. The authors are very grateful to Transpharmation Ltd. for supporting this work. Many thanks to Mrs Carol Hoy Ncert A&CC, VTS (Anesthesia/Analgesia), RVN for the figure provided. Thank you to all the dog owners for participating in this study.

Authors' contributions

SSM: design, data management, data interpretation, statistical analysis and preparation of manuscript; YC: data interpretation, statistical analysis and preparation of manuscript; SA: data interpretation, statistical analysis and preparation of manuscript; AF: data interpretation and preparation of manuscript; HAV: data interpretation and preparation of manuscript; LP: design, data management, data interpretation, statistical analysis and preparation of manuscript.

References

Andersen KG, Kehlet H, Aasvang EK (2014) Test-retest Agreement and Reliability of Quantitative Sensory Testing 1 Year after Breast Cancer Surgery. Clin J Pain 31, 393-403.

Backonja MM, Attal N, Baron R et al. (2013) Value of quantitative sensory testing in neurological and pain disorders: NeuPSIG consensus. Pain 154**,** 1807-1819.

Baron R, Dickenson AH (2014) Neuropathic pain: precise sensory profiling improves treatment and calls for back-translation. Pain 155, 2215-2217.

Blankenburg M, Boekens H, Hechler T et al. (2010) Reference values for quantitative sensory testing in children and adolescents: Developmental and gender differences of somatosensory perception. Pain 149**,** 76-88.

Briley JD, Williams MD, Freire M et al. (2014) Feasibility and repeatability of cold and mechanical quantitative sensory testing in normal dogs. Vet J 199**,** 245-250.

Brown DC, Boston RC, Coyne JC et al. (2008) Ability of the canine brief pain inventory to detect response to treatment in dogs with osteoarthritis. J AmVet Med Assoc 233**,** 1278-1283.

Brown DC (2012) Quantitative sensory testing: A stimulating look at chronic pain. The Veterinary Journal 193, 315-316.

Brydges NM, Argyle DJ, Mosley JR et al. (2012) Clinical assessments of increased sensory sensitivity in dogs with cranial cruciate ligament rupture. Vet J 193**,** 545-550.

Chaplan SR, Bach FW, Pogrel JW et al. (1994) Quantitative assessment of tactile allodynia in the rat paw. J Neurosci Methods 53**,** 55-63.

Chong PS, Cros DP (2004) Technology literature review: quantitative sensory testing. Muscle Nerve 29**,** 734-747.

455 Coleman KD, Schmiedt CW, Kirkby KA et al. (2014) Learning Confounds Algometric
456 Assessment of Mechanical Thresholds in Normal Dogs. Vet Surg 43**,** 361-367.
457 Edwards RR, Dworkin RH, Turk DC et al. (2016) Patient phenotyping in clinical trials of
458 chronic pain treatments: IMMPACT recommendations. Pain May 5. Epub ahead of
459 print.
460 Finnerup NB, Haroutounian S, Kamerman P et al. (2016) Neuropathic pain: An updated
461 grading system for research and clinical practice. Pain. 2016 Apr 28. Epub ahead of
462 print.
463 Gorney AM, Blau SR, Dohse CS et al. (2016) Mechanical and Thermal Sensory Testing in
464 Normal Chondrodystrophoid Dogs and Dogs with Spinal Cord Injury caused by
465 Thoracolumbar Intervertebral Disc Herniations. J Vet Intern Med.;30,627-35. doi:
466 10.1111/jvim.13913.
467 Greenspan JD (2001) Quantitative assessment of neuropathic pain. Curr Pain Headache Rep
468 5**,** 107-113.
469 Grone E, Crispin A, Fleckenstein J et al. (2012) Test order of quantitative sensory testing
470 facilitates mechanical hyperalgesia in healthy volunteers. J Pain 13**,** 73-80.
471 Hansson P (2002) Neuropathic pain: clinical characteristics and diagnostic workup. Eur J
472 Pain 6, Supplement A**,** 47-50.
473 Hansson P, Backonja M, Bouhassira D (2007) Usefulness and limitations of quantitative
474 sensory testing: clinical and research application in neuropathic pain states. Pain 129**,**
475 256-259.
476 Harris LK, Murrell JC, van Klink EG et al. (2015) Influence of experimental protocol on
477 response rate and repeatability of mechanical threshold testing in dogs. Vet J.

478 Hoffmann MV, Kastner SB, Kietzmann M et al. (2012) Contact heat thermal threshold
479 testing in beagle dogs: baseline reproducibility and the effect of acepromazine,
480 levomethadone and fenpipramide. BMC Vet Res 8**,** 206.
481 Hunt JR, Grint NJ, Taylor PM et al. (2013) Sedative and analgesic effects of buprenorphine,
482 combined with either acepromazine or dexmedetomidine, for premedication prior to
483 elective surgery in cats and dogs. Vet Anaesth Analg 40**,** 297-307.
484 Knazovicky D, Helgeson ES, Case B et al. (2016) Widespread somatosensory sensitivity in
485 naturally occurring canine model of osteoarthritis. Pain157,1325-32. doi:
486 10.1097/j.pain.0000000000000521.
487 Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data.
488 Biometrics 33**,** 159-174.
489 Lascelles BDX (2013) Getting a sense of sensations. Vet J 197**,** 115-117.
490 Lee KM, Lee J, Chung CY et al. (2012) Pitfalls and important issues in testing reliability
491 using intraclass correlation coefficients in orthopaedic research. Clinics in orthopedic
492 surgery 4, 149-155.
493 Mathews KA (2008) Neuropathic pain in dogs and cats: if only they could tell us if they hurt.
494 Vet Clin Small Anim 38, 1365-1414.
495 Moloney NA, Hall TM, Doody CM (2012) Reliability of thermal quantitative sensory testing:
496 a systematic review. J Rehabil Res Dev 49**,** 191-207.
497 Moore SA, Hettlich BF, Waln A (2013) The use of an electronic von Frey device for
498 evaluation of sensory threshold in neurologically normal dogs and those with acute
499 spinal cord injury. Vet J 197**,** 216-219.
500 Reimer M, Helfert SM, Baron R (2014) Phenotyping neuropathic pain patients: implications
501 for individual therapy and clinical trials. Curr Opin Support Palliat Care 8, 124-129.

Rolke R, Baron R, Maier C et al. (2006) Quantitative sensory testing in the German Research Network on Neuropathic Pain (DFNS): Standardized protocol and reference values. Pain 123**,** 231-243.

Song RB, Basso DM, da Costa RC et al. (2016) von Frey anesthesiometry to assess sensory impairment after acute spinal cord injury caused by thoracolumbar intervertebral disc extrusion in dogs. Vet J 209, 144-149.

Taylor PM, Crosignani N, Lopes C et al. (2015a) Mechanical nociceptive thresholds using four probe configurations in horses. Vet Anaesth Analg. doi: 10.1111/vaa.12274. [Epub ahead of print]

Taylor PM, Hoare HR, de Vries A et al. (2015b) A multicentre, prospective, randomised, blinded clinical trial to compare some perioperative effects of buprenorphine or butorphanol premedication before equine elective general anaesthesia and surgery. Equine Vet J. doi: 10.1111/evj.12442. [Epub ahead of print]

Taylor PM, Robertson SA, Dixon MJ (2007) Evaluation of the use of thermal thresholds to investigate NSAID analgesia in a model of inflammatory pain in cats. J Feline Med Surg 9**,** 313-318.

Tomas A, Marcellin-Little DJ, Roe SC et al. (2014) Relationship Between Mechanical Thresholds and Limb Use in Dogs With Coxofemoral Joint OA-Associated Pain and the Modulating Effects of Pain Alleviation From Total Hip Replacement on Mechanical Thresholds. Vet Surg 43**,** 542-548.

Vangeneugden T, Laenen A, Geys H et al. (2004) Applying linear mixed models to estimate reliability in clinical trial data with repeated measurements. Control Clin Trials 25**,** 13-30.

525 Walk D, Sehgal N, Moeller-Bertram T et al. (2009) Quantitative sensory testing and
526 mapping: a review of nonautomated quantitative methods for examination of the
527 patient with neuropathic pain. Clin J Pain 25**,** 632-640.
528 Werner MU, Petersen MA, Bischoff JM (2013) Test-retest studies in quantitative sensory
529 testing: a critical review. Acta Anaesthesiol Scand 57**,** 957-963.
530 Whitaker LH, Reid J, Choa A et al. (2016) An Exploratory Study into Objective and
531 Reported Characteristics of Neuropathic Pain in Women with Chronic Pelvic Pain.
532 PloS one 11, e0151950.
533 Williams MD, Kirkpatrick AE, Griffith E et al. (2014) Feasibility and repeatability of thermal
534 quantitative sensory testing in normal dogs and dogs with hind limb osteoarthritis-
535 associated pain. Vet J 199, 63-7.
536 Woolf CJ, Salter MW (2000) Neuronal plasticity: increasing the gain in pain. Science 288**,**
537 1765-1769.
538 Younis S, Maarbjerg S, Reimer M et al. (2016) Quantitative sensory testing in classical
539 trigeminal neuralgia-a blinded study in patients with and without concomitant
540 persistent pain. Pain 157, 1407-1414.
541

List of figures:

Figure 1 Consort flow diagram of dogs included in the study. TST_{UD}: *up-down technique* method of testing*;* $TST_{50\%}$: *50% of response technique* method of testing; MT: mechanical threshold 2, 4 and 8 mm size probe; HT: heat threshold; CT: cold threshold.

Figure 2 Body areas tested and anatomical localization.

1) Left and right tibias: mid-point between the stifle joint and the hock on the lateral aspect of the tibia;

2) Left and right humeri: mid-point between the scapulo-humeral joint and the elbow on the lateral aspect of the humerus;

3) Left and right neck: mid-point between the atlas wings and the cranial aspect of the body of the scapula on the lateral aspect of the neck;

4) Left and right thoraco-lumbar (T-L): palpate the last rib-vertebrae union. At that level, palpate the spinous process. Testing point is located 1cm (small dog) to 3cm (large dog) lateral to the spinal process;

5) Left abdomen: mid-point between midline and the fold of the flank.

Illustration courtesy of Mrs Carol Hoy

Figure 3 Median, interquartile range and min-max thresholds of the sensory threshold examination protocol (STEP). The three different weight categories are displayed on the figure. For statistical difference between body areas see Table S3.

A, B, C: MT: mechanical threshold with the 2, 4, and 8 mm size probe; N: Newton

D, E: TST: tactile sensitivity threshold; gf: grams of force;

F: HT: heat thresholds (^{0}C)

G CT: cold latency (seconds),

H: Probability of response to von Frey filaments (TST) on different body areas (%).TST_{UD}: *up-down technique* method of testing; $TST_{50\%}$: *50% of response technique* method of testing.

I: Probability of response to Cold stimulus (%)

Table 1 Mechanical (MT) and heat thresholds mean, standard deviation (SD) and range obtained for the different probes and the different weight categories. Response to tactile stimulus and cold stimulus (%), tactile sensitivity threshold (TST) method 1 and 2 and cold latency (at 0^0C), median and range obtained in the different body areas and weight categories. Values were log-transformed for the analysis and back-transformed for MT.

Variable	Dog size	Body Area				
		Tibia	Humerus	Neck	T-L	Abdomen
MT 2 mm probe (N)	Small (1 −8 kg)	4.6 ± 1.6 (1.7−10.50)	4.3 ± 1.5 (1.62 - 9.12)	7.9 ± 1.3 (5.13 - 11.75)	5.8 ± 1.6 (1.95 - 12.02)	3.4 ± 1.6 (1.74 - 6.76)
	Medium (9 - 22kg)	5.6 ± 1.4 (3.63 - 13.18)	5.6 ± 1.4 (2.69 - 12.02)	9.8 ± 1.3 (5.25 - 15.49)	5.9 ± 1.5 (2.29 - 14.79)	2.8 ± 1.6 (1.55 - 5.25)
	Large (23-40kg)	7.1 ± 1.6 (2.51 - 18.62)	7.1 ± 1.4 (3.89 - 14.79)	13.5 ± 1.5 (3.39 - 25.12)	8.3 ± 1.5 (2.69 - 20.42)	4.8 ± 1.7 (1.05 - 11.22)
MT 4 mm probe (N)	Small (1−8 kg)	6.5 ± 1.6 (2.45 - 14.79)	5.7 ± 1.5 (1.41 - 15.14)	9.8 ± 1.4 (3.89 - 16.98)	8.1 ± 1.5 (3.09 - 17.38)	4.4 ± 1.7 (1.86 - 8.32)
	Medium (9- 22kg)	8.3 ± 1.5	7.3 ± 1.3	11.9 ± 1.3	8 ± 1.4	4.3 ± 1.6

		(2.63 - 14.45)	(4.68 - 11.75)	(8.13 - 19.50)	(3.72 - 15.140	(1.86 - 10.47)
	Large (23-40kg)	9.5 ± 1.4 (3.8 - 20.89)	9.9 ± 1.3 (3.72 -16.98)	16.1 ± 1.3 (7.94 - 22.91)	10.3 ± 1.5 (3.31 - 24.55)	7.2 ± 1.5 (3.09 - 15.49)
MT 8 mm probe (N)	Small (1−8 kg)	9.7 ± 1.3 (5.89 - 16.22)	8.9 ± 1.4 (2.69 - 15.49)	12.2 ± 1.4 (4.47 - 19.95)	12.9 ± 1.4 (6.31 - 21.88)	7.2 ± 1.6 (2.63 - 11.75)
	Medium (9- 22kg)	11.1 ± 1.3 (7.24 - 19.05)	11.1 ± 1.4 (5.37 - 18.20)	15.9 ± 1.3 (9.77 - 22.91)	11.4 ± 1.5 (2.75 - 22.39)	6 ± 1.5 (2.75 - 2.75)
	Large (23-40kg)	13.5 ± 1.4 (6.31 - 24.55)	13.8 ± 1.3 (7.41 - 25.12)	20.6 ± 1.4 (7.41 - 34.67)	15 ± 1.5 (4.47 - 33.88)	9.8 ± 1.7 (3.8 - 29.51)
Heat Threshold (°C)	Small (1−8 kg)	43.0 ± 2.5 (39.10 - 50.25)	45.0 ± 3 (40.30 - 50.87)	48.2 ± 3.2 (44.10 - 55.00)	47.5 ± 3.5 (42.23 - 55.00)	44.7 ± 3.3 (40.90 - 40.90)
	Medium (9- 22kg)	43.8 ± 3.1 (39.57 - 50.20)	46.6 ± 3.6 (41.40 - 55.00)	48.5 ± 3.9 (40.70 - 55.00)	47.3 ± 3.5 (40.60 - 55.00)	43.9 ± 2.1 (40.70 - 46.50)
	Large (23-40kg)	46.4 ± 4 (38.80 - 55.00)	49.4 ± 3.7 (39.85 - 55.00)	51.9 ± 3.4 (40.00 - 55.00)	51.3 ± 3.4 (43.27 - 55.00)	46.8 ± 4.5 (37.75 - 55.00)
TST_{UD}	Small (1 - 8 kg)	(22/32) 68.7 %	(22/32) 68.7%	(14/32) 43.7%	(22/32) 68.7%	(8/16) 6.3%

(gf)		79.43 (7.84 - 597.50)	130.80 (8.88 - 597.50)	597.50 (24.05 - 597.50)	164.40 (11.91 - 597.50)	372 (11.91 - 597.50)
	Medium (9 - 22kg)	(14/18) 77.7% 180 (46.64 -597.50)	(6/18) 33.3% 597.50 (72.21 - 597.50)	(2/18) 11.1% 597.50 (279.1 - 597.50)	(11/18) 61.1% 311.70 (101.2 - 597.50)	(5/9) 55.5% 597.50 (71.21 - 597.50)
	Large (23 - 40kg)	(24/44) 54.5% 311.70 (6.82 - 597.50)	(13/44) 29.5% 597.50 (7.55 - 597.50)	(5/44) 11.3% 597.50 (47.66 - 597.50)	(23/44) 52.3% 303.10 (7.94 - 597.50)	(9/22) 40.9% 597.50 (11.66 - 597.50)
$TST_{50\%}$ (gf)	Small (1−8 kg)	(11/16) 68.7 % 300 (180 - 300)	(6/16) 37.5% 300 (100 - 300)	(2/16) 12.5% 300 (180 - 300)	(9/16) 56.2% 300 (180 - 300)	(2/7) 28.5% 300 (300 - 300)
	Medium (9- 22kg)	(10/14) 71.42% 100 (4 - 300)	(15/18) 83.3% 240 (4 - 300)	(6/18) 33.3% 300 (180 - 300)	(17/18) 94.4% 180 (8 - 300)	(5/7) 71.4% 180 (4 - 300)
	Large (23-40kg)	(4/4) 100% 37.5	(2/4) 50% 300	(2/4) 50% 300	(4/4) 100% 300	(2/2) 100% 300

		(15 - 300)	(300 - 300)	(300 - 300)	(300 - 300)	(300 - 300)
Cold ^{0}C (seconds)	Small (1−8 kg)	(43/123) 35 % 60 (11.41 - 60)	(18/123) 14.6% 60 (28.17 - 60)	(18/126) 14.6% 60 (31.40 - 60)	(15/126) 11.9% 60 (11.97 - 60)	(4/63) 6.3% 60 (48.33 - 60)
	Medium (9- 22kg)	(79/78) 35.2% 58.84 (9.83 - 60)	(19/75) 25.3% 60 (9.40 - 60)	(5/78) 6.4% 60 (32.40 - 60)	(8/78) 10.2% 60 (9.30 - 60)	(10/39) 25.6% 60 (21.8 - 60)
	Large (23-40kg)	(71/132) 53.8% 43.25 (18.50 - 60)	(38/129) 30.2% 60 (41 - 60)	(38/132) 28.7% 56.36 (6.4 - 60)	(35/129) 27.1% 60 (9.38 - 60)	(24/66) 36.6% 54.16 (12.27 - 60)

T- L, thoraco- lumbar area; (gf), gram of force; N, Newton; TST_{UD}, tactile sensitivity thresholds *up-down technique* method; $TST_{50\%}$, tactile sensitivity thresholds *50% response technique* method

Table 2 Results of linear mixed model and general linear mixed model. Effect of body area, weight category, age category, sex and factors of reliability and performance of the protocol (occasion, feasibility scores and period of testing) on TST, MT, CT, HT.

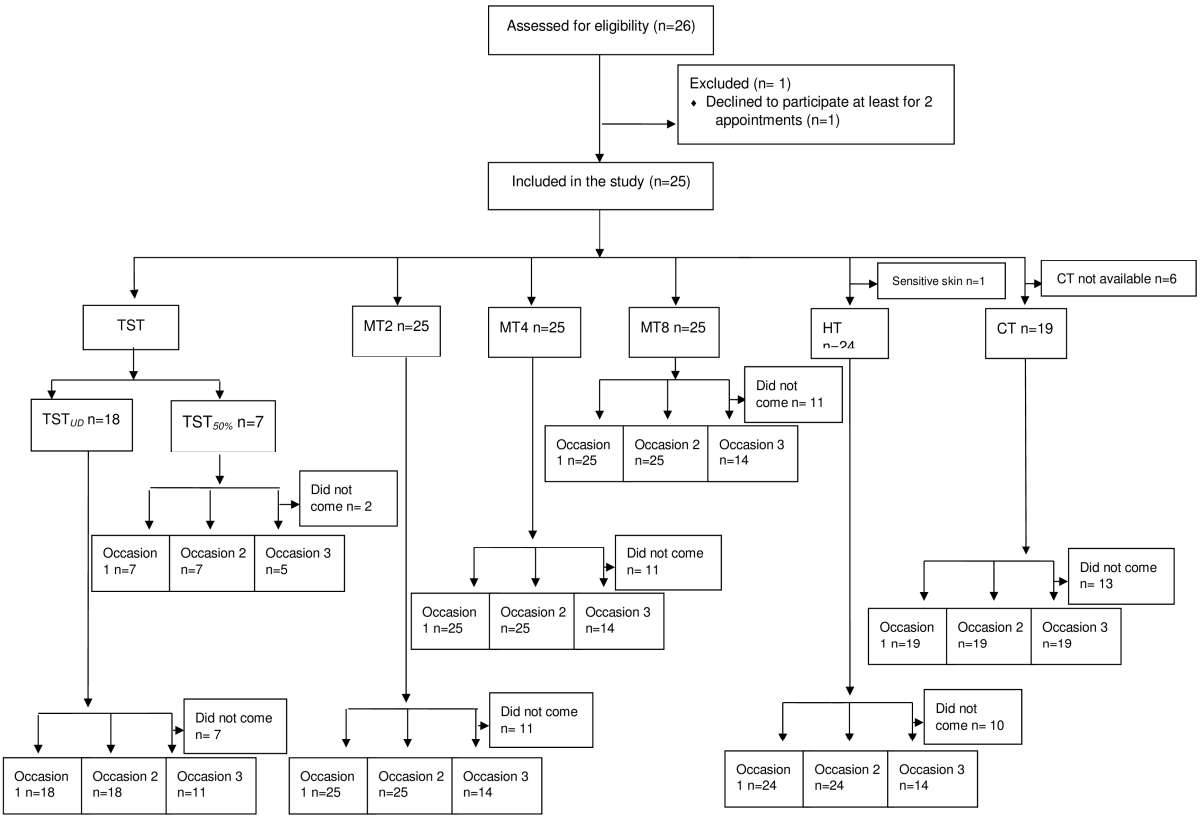
Fixed Effect	**TST_{UD}**	**$TST_{50\%}$**	**MT2mm**	**M4mm**	**MT8mm**	**HT**	**CT**
BA	0.783	0.001*	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*
L/R side	0.642	0.478	0.685	0.405	0.760	0.884	0.515
Weight category	0.06	<0.001*	<0.001*	<0.001*	<0.001*	0.008*	0.037*
Age category	0.076	0.408	0.145	0.384	0.846	0.041*	0.448
Sex	0.006*	0.009*	0.009*	0.131	0.032*	0.021*	0.088
Skin temperature	-	-	-	-	-	0.457	0.082
Room temperature	-	-	-	-	-	0.365	0.087
Feasibility score	0.004*	0.060	0.557	0.144	0.852	0.08	0.221
Occasion (1,2,3)	0.825	0.119	0.747	0.470	0.158	0.930	0.004*
Period of testing	0.573	-	0.050*	0.043*	0.014*	0.934	0.067

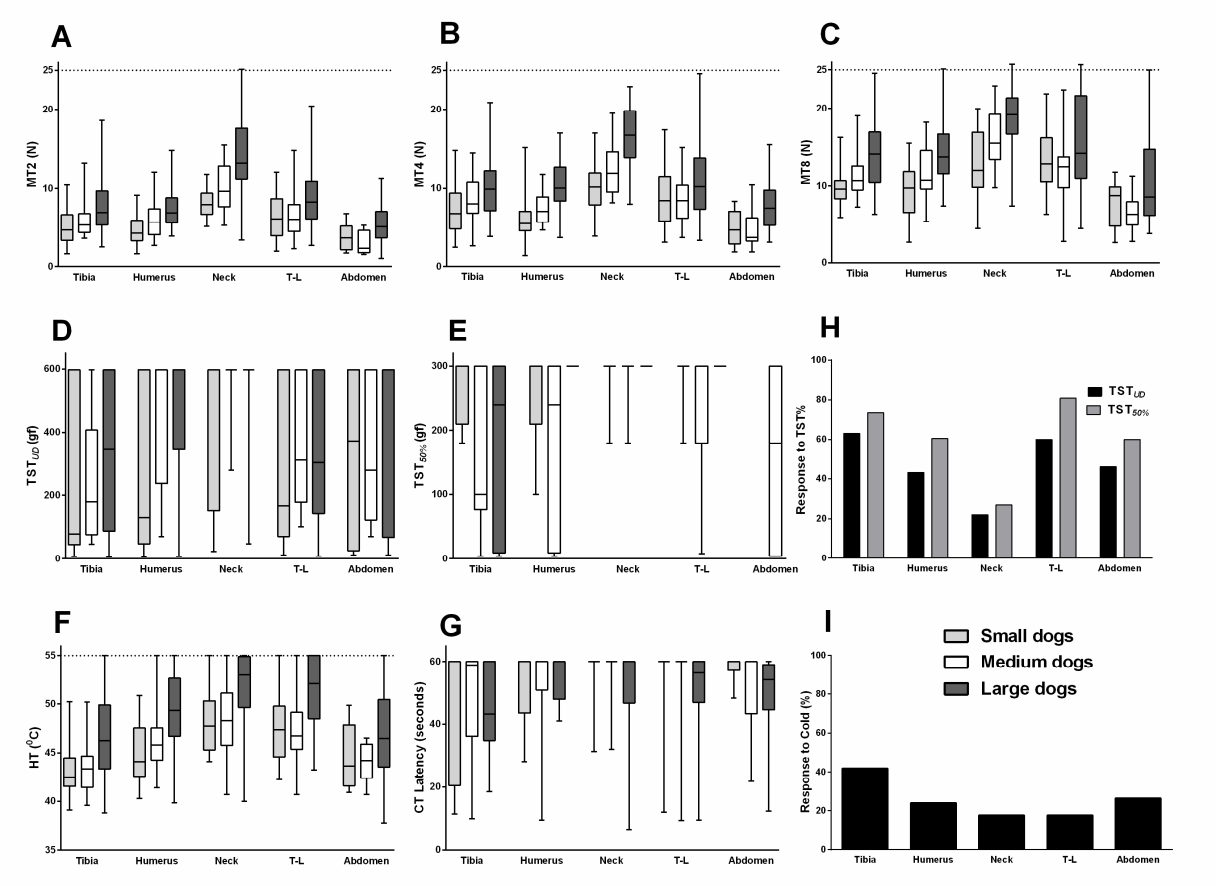
BA, body areas; L/R, left/right side; TST_{UD}, tactile sensitivity thresholds *up-down technique* method; $TST_{50\%}$, tactile sensitivity thresholds *50% response technique* method; MT, mechanical thresholds; HT, heat thresholds; CT cold latency thresholds. $P < 0.005$

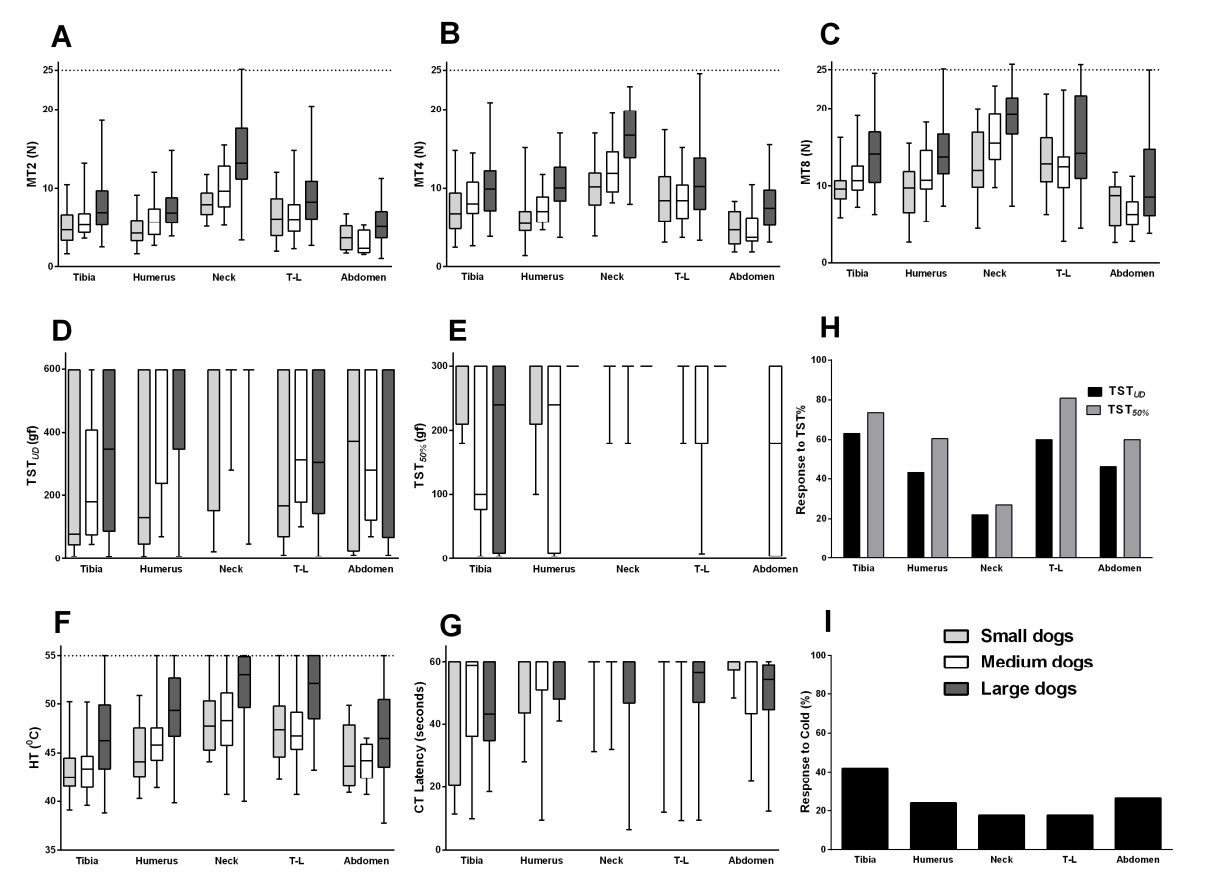
Table 3 Intra-class correlation coefficient (ICC) and 95% confidence interval (CI) of the different tests of the STEP and ICC of MT for the two different periods of testing, where differences in MT were observed in the linear mixed effect model. There is a mild improvement in ICCs between period 1 and period 2 with the 3 different probes.

	TST_{UD}	**$TST_{50\%}$**	**MT2**	**MT4**	**MT8**	**HT**	**CT**
ICC	0.001	0.71	0.72	0.69	0.68	0.58	0.51
95% CI	N/A	0.1-1	0.58-0.86	0.52-0.85	0.51-0.84	0.34-0.86	0.22-0.77
Period 1 ICC	N/A	N/A	0.72	0.65	0.65	N/A	N/A
Period 2 ICC	N/A	N/A	0.75	0.78	0.76	N/A	N/A

TST_{UD}, tactile sensitivity thresholds *up-down technique* method; $TST_{50\%}$, tactile sensitivity thresholds *50% response technique* method; MT, mechanical thresholds; HT, heat thresholds; CT cold latency thresholds.




A
MT2 (N)
B
MT4 (N)
C
MT8 (N)
Tibia
Humerus
Neck
T-L
Abdomen
D
TST$_{UD}$ (gf)
E
TST$_{50\%}$ (gf)
H
Response to TST%
TST$_{UD}$
TST$_{50\%}$
F
HT (^{0}C)
G
CT Latency (seconds)
I
Response to Cold (%)
Small dogs
Medium dogs
Large dogs


A
MT2 (N)
B
MT4 (N)
C
MT8 (N)
Tibia
Humerus
Neck
T-L
Abdomen
D
TST$_{UD}$ (gf)
E
TST$_{50\%}$ (gf)
H
Response to TST%
TST$_{UD}$
TST$_{50\%}$
F
HT (^{0}C)
G
CT Latency (seconds)
I
Response to Cold (%)
Small dogs
Medium dogs
Large dogs

Appendix 1 Feasibility scores. Adapted from (Briley et al. 2014)

Feasibility score	Description
0 – No problem	Minimum restraint needed; excellent cooperation; clear reaction to stimuli
1 – Mild difficulty	Mild restraint needed; good cooperation; clear reaction to stimuli
2 – Moderate difficulty	Moderate restraint needed; good cooperation >50% of the time; mild sensitivity to being touched; mild variation in reaction to stimuli
3 – Significant difficulty	Significant restraint needed and resisted sternal position; good cooperation <25% of the time; moderate sensitivity to being touched; moderate variation in reaction to stimuli
4 – Extreme difficulty	Constant restraint required; not cooperative; unclear reaction to stimuli, not confident in data collected
5 – Impossible	Could not collect data due to the dog's disposition and/or lack of confidence in the reactions seen being due to the stimulus

Appendix 2 Body weight and age categories of the sample of dogs.

	Dogs		
Category	Classification	*n*	%
Age (Years)	Young (0.3 – 3)	9	36
	Adult (4 – 6)	9	36
	Senior (> 6)	7	28
Weight (kg)	Small (1 -8)	10	40
	Medium (9 -22)	6	24
	Large (23-40)	9	36

n, number of dogs

Appendix 3 Post Hoc comparisons, odds ratio (OR) and estimated mean differences comparing body areas, weight category, age category, sex, feasibility score, occasion tested and period of testing. Main differences for mechanical thresholds (MT) 2, MT4 and MT8 are displayed as back log transformed.

Pairwise comparison	**$pTST_{UD}$**		**$pTST_{50\%}$**		**MT2 (N)**		**MT4 (N)**		**MT8 (N)**		**HT (^{0}C)**		**pCT**	
BA	OR	P-value	OR	P-value	Mean difference	P-value	Mean difference	P-value	Mean difference	P-value	Mean difference	P-value	OR	P-value
Tibia - Humerus	-	-	0.3	0.105	1.0	0.429	1.1	0.140	1.01	0.697	-2.7	<0.001*	0.7	<0.001*
Tibia - Neck	-	-	0.0	<0.001*	-1.7	<0.001*	-1.6	<0.001*	-1.4	<0.001*	-5.4	<0.001*	0.7	<0.001*
Tibia - T-L	-	-	1.1	0.936	-1.1	0.003*	-1.1	0.023*	-0.3	0.001*	-4.5	<0.001*	0.7	<0.001*
Tibia - Abdomen	-	-	0.4	0.304	1.6	<0.001*	1.5	<0.001*	1.5	<0.001*	-0.9	0.023*	0.7	<0.001*
Humerus -. neck	-	-	0.1	0.008*	-1.8	<0.001*	-1.6	<0.001*	-1.4	<0.001*	-2.7	<0.001*	0.9	0.028*
Humerus - T-L	-	-	3.7	0.083	-1.1	<0.001*	-1.2	<0.001*	-1.1	<0.001*	-1.7	<0.001*	0.9	0.032*

Humerus - Abdomen	-	-	1.3	0.768	1.5	<0.001*	1.4	<0.001*	1.5	<0.001*	1.7	0.001*	1.0	0.642
Neck - T-L	-	-	24.5	<0.001*	1.5	<0.001*	1.4	<0.001*	1.2	<0.001*	0.9	0.014*	1.0	0.850
Neck - Abdomen	-	-	9.0	0.024*	2.8	<0.001*	2.4	<0.001*	2.1	<0.001*	4.4	<0.001*	1.0	0.084
T-L - Abdomen	-	-	0.3	0.269	1.8	<0.001*	1.7	<0.001*	1.7	<0.001*	3.5	<0.001*	1.1	0.095
Weight category														
Small - Medium	-	-	0.03	<0.001*	-1.4	0.008*	-1.3	0.029*	-1.4	0.012*	-1.1	0.316	1.2	0.063
Small – Large	-	-	0.0000008	0.001*	-1.9	<0.001*	-1.6	<0.001*	-1.8	<0.001*	-0.3	0.001*	1.5	0.010*
Medium - Large	-	-	0.00002	0.006*	-1.3	0.010*	-1.3	0.005*	-1.3	0.014*	-2.6	0.020*	1.1	0.485
Age category														

Young - Adults	-	-	-	-	-	-	-	-	-	-	2.2	0.035*	-	-
Young - Senior	-	-	-	-	-	-	-	-	-	-	-0.6	0.562	-	-
Adults - Senior	-	-	-	-	-	-	-	-	-	-	-2.9	0.013*	-	-
Sex														
Female - Male	0.5	0.006*	0.04	0.009*	1.3	0.009*	-	-	1.2	0.023*	2.5	0.021*	-	-
Feasibility score														
0-1	0.3	0.001*	-	-	-	-	-	-	-	-	-1.2	0.296	-	-
0-2	0.4	0.028*	-	-	-	-	-	-	-	-	1.5	0.153	-	-
1-2	1.3	0.214	-	-	-	-	-	-	-	-	2.7	0.013*	-	-
Occasion														
1 - 2	-	-	-	-	-	-	-	-	-	-	-	-	1.1	0.223
1 - 3	-	-	-	-	-	-	-	-	-	-	-	-	0.9	0.006*

2 - 3	-	-	-	-	-	-	-	-	-	-	-	-	0.8	0.003*
Period of testing														
1-2	-	-	-	-	-1.2	0.050*	-1.2	0.043*	-9.1	0.014*	-	-	-	-

BA, body area; $pTST_{UD}$, response to tactile sensitivity *up-down technique* method and $pTST_{50\%}$ with *50% response technique* method (any of the von Frey filaments) ; MT, mechanical thresholds; HT: heat thresholds; pCT: response to 0^0 C before 60 seconds; N: newton; T- L: thoraco-lumbar area; (-), no significant difference for covariate/cofactor on this test; $P < 0.05$