Journal of Novel Physiotherapies

Georgoudis, et al, J Nov Physiother 2014, 4:5
http://dx.doi.org/10.4172/2165-7025.1000234

Research Article Open Access

Reliability Measures of Subcutaneous Pressure Pain Threshold Measurements: A Proposed Method of Assessing Painful Musculoskeletal Disorders

George Georgoudis[1*], Jacqueline Oldham[2], Paul J Watson[3] and Eirini Grammatopoulou[1]

[1]*School of Physiotherapy, Technological Educational Institution of Athens, Athens, Greece*

[2]*Graduate School of Medicine, Manchester, UK*

[3]*University of Leicester, College of Medicine, Biological Sciences, and Psychology, UK*

[*]**Corresponding author:** George Georgoudis, School of Physiotherapy, Technological Educational Institution of Athens, Athens, Greece, Tel: 302104014920; E-mail: gg@hol.gr

Received date: October 15, 2014; **Accepted date:** November 20, 2014; **Published date:** November 27, 2014

Abstract

Background: Pressure algometry is a widely used method aiming to investigate deep tissue mechanical pain threshold in painful musculoskeletal disorders. Considering from the literature the input of skin and subcutaneous tissue in the determination of pressure pain threshold, and the distinctive role in clinical setting of the subcutaneous pressure threshold measurement (sPPT), it becomes necessary to investigate the reliability of the sPPT measurements. This study examines the short and longer-term intra-examiner reliability of sPPT measurements.

Methods: Thirty subjects were examined at 3 pre-determined time intervals, in various parts of the body, by the same examiner in the same clinical setting. Short (between days) and longer-term (after a week) test-retest reliability was investigated. A number of statistical procedures was computed: ICC, ANOVA, standard error of measurement (SEM), smallest detectable difference (SDD) and confirmatory scatterplots.

Results: The ICC statistic reached good to excellent levels for both types of reliability (0.70 – 0.97). ANOVA produced non-significant differences at all measurements. SEM gave a satisfactory overall mean value of 6.07 kPa (short-term) for all measuring sites (range: 3.2-10.1) and similarly for the longer-term reliability (Mean=6.2, range: 3.1-11.2). SDD were satisfactory in the majority of the measurements [short: Mean=11.76, range: 5.9-20, longer: Mean=12.0, range: 6.5-22] but relatively high in some measuring sites, nevertheless fully acceptable and safely within the pressure algometry limits as defined in the literature.

Conclusions: Subcutaneous pressure threshold measurement is a reliable to reproduce with stable results procedure, either for short or longer periods of time (from 1 to 7 days), across the whole body (upper – lower body), by the same examiner. Its fluctuation in absolute values is within the literature limits of deep pressure algometry.

Keywords: Subcutaneous pressure pain threshold; Pressure pain threshold; Pressure algometry; Reliability

Introduction

Pressure pain threshold measurements (PPT) are nowdays used routinely in clinical practice to assess an extended number of pathologies including musculoskeletal conditions in humans [1] and animals [2]. However, reliance on PPT as a measure of deep tissue tenderness disregards the role of superficial tissues (skin and subcutaneous tissues), which inevitably participate in the process [3].

Anatomically, pressure algometry stimulates the cutaneous Aβ afferents and the second-order neurons in the nucleus caudalis that respond to stimulation of deep tissue [4]. However, these neurons also receive converging input from the skin in a systematic manner [4]. This neuroanatomic evidence is supported from clinical data, where it has been shown that treatments applied topically to the skin can result in reduction of deeper pain, tenderness (e.g. deep PPT) and/or the subjective feeling of pain intensity, through the activation of the central nervous system (dorsal horn level and above). For example, the transient application of cold spray and stretching [5], iontophoresis [6], ultrasound [7], transcutaneous electrical nerve stimulation (TENS) [8], intradermally infiltrating the area with a local anesthetic [9], superficial needling [10] are clinical evidence that highlight the role of superficial tissues in alleviating pain. Therefore, superficial skin and subcutaneous tissues may indeed play a role in the determination of deeper tissue PPT values as already been mentioned by a few studies [3,11,12]. The above findings suggest that the overlying skin/subcutis contributes to quantitative assessment of deep tissue PPT which may be influenced by mechanical cutaneous sensitivity.

Greater support for this model can be argued with the findings of previously published studies based on electrically induced pain threshold since the electric and mechanical PT are linearly correlated [13]. According to Vecchiet's research team, the three parietal tissues (skin, subcutaneous tissue, muscle) have a different electric pain threshold which is progressively higher for the superficial than the deep structures [14]. This electric pain threshold is modified for all three tissues in relation to the intensity of pain felt [15], as is the case for PPT and sPPT [16]. Furthermore, in a clinical situation where there is a predominant muscle tissue pathology – i.e. a MTrP, it was noted that pain threshold (electric and mechanical) from all three parietal tissues was modified, even in the pre-clinical phase [15,16].

Thus, the summative nature of PPT, regarding the pain threshold properties of all three parietal tissues has been demonstrated. This imposes the necessity to record skin and subcutaneous pain threshold in addition to the deep tissues' pain threshold, as it has been advocated by a number of authors [3,17,18].

Unfortunately, research to date, has focused mainly on the determination of skin and subcutaneous pain threshold via electric current [13-15,19]. The great disadvantage of determining the electric pain threshold is its invasive character (needle electrodes) and is therefore impractical to implement in a routine daily clinical environment. On the other hand, the major benefits with pressure pain threshold determination are that it is possible to measure the applied pressure (estimate of tenderness) precisely and on a ratio scale, with a non-invasive technique, in a time-efficient way, without the use of complex equipment [20]. The easiness of the application of the technique in the clinical environment is hoped to render the subcutaneous pressure algometry a favorite method.

The methodology to perform a subcutaneous pressure pain threshold measurement has been described by Fischer: "…a skinfold is produced between the examiner's thumb and the tip of the algometer which is pressed against the thumb, applying pressure to the subcutaneous tissues within the skinfold. The pressure is increased at a continuous rate of 1 kg/sec, in a manner similar to the technique used for measurement of deep tissue tenderness … it is suggested that the criterion of abnormality applied in deep tissues be utilized for subcutaneous measurements as well…" [17] (pg 22). However, the standards suggested (technical parameters) and the criteria adopted are self-quoted by the author based on what is valid for the deep PPT measurement. Considering the neuro-anatomic differences – e.g. the density and number of mechano-receptors [21], the variability in sensitivity and other fundamental variations of the three parietal tissues (e.g. physical properties of the tissues, the tipsize of the algometer, etc), it is doubted if the same criteria can be applied. The most appropriate technical parameters to measure sPPT need to be standardized in future through published evidence. In the meantime, researchers may purposefully attempt the standards (technical parameters) and criteria that seem best, based on their clinical experience.

Reports of estimated sPPT values come from studies from the Vecchiet research team [13-16,19] and Fischer's review on pressure algometry [17]. Unfortunately, no study in a consistent, clear manner reports the methodology that resulted in the recorded sPPT values. Sometimes, the actual sPPT values are not even mentioned, instead a correlation coefficient value with the electric pain threshold measurement is reported. No data on the technical parameters of the algometer used are quoted; neither the exact locations of measuring sPPT are mentioned (e.g. above a bony, muscular, tendinous area). Therefore, although Fischer concluding his review stated that: "… regardless of the criterion of abnormality, quantified subcutaneous tissue pressure pain sensitivity can be used for evaluation of treatment results" [17] p23, it is vital first to examine the reliability of the method before implementing sPPT in clinical practice.

It is the objective of this study to demonstrate the test-retest reliability of sPPT measurements using pre-defined assessment criteria. Specifically, the aim is to investigate the short and longer-term intra-examiner reliability of sPPT measurements.

Methods

Sampling

A randomised cluster sampling technique was employed to establish the sample of the study. A total of 30 (17 males) subjects attending the Kinesiology module in the Technological Educational Institute of Lamia, Greece, were recruited. Sampling was based on a random numbers selection (generated by a computer software application) from the total population of students attending the modules (approximately 120). Only one subject was excluded from the study due to excessive levels of depression and anxiety (see exclusion criteria). Demographics of the sample are presented in Results.

Literature suggests 30-35 subjects for reliability designs in biomedical research when α=0.05 and β=0.20 [22]. On these lines a pragmatic sample of 30 subjects was included fulfilling the criteria. Specifically, the exclusion criteria selected for this study were based on findings concerning PPT rather than sPPT studies, since there are no sPPT related studies in the literature. The specific criteria are detailed below:

Pain complaints and/or treatment at the shoulder/neck area within the last 3 months (as self-reported by the examinees).

Concurrent signs of malaise or fever, involuntary weight loss, rheumatic disease, cancer or psychiatric history (as self-reported by the examinees).

Clinically apparent major depression [23], (As assessed by the HADs scale and the clinical opinion of the examiner). One female was excluded due to this criterion.

Known endocrine or metabolic disorders which are not controlled by the given medical treatment.

Severely hypotensive or hypertensive (blood pressure) individuals.

Previous recent (within 6 months) open wound in the cervical spine or upper extremity (as self-reported by the examinees), only on the basis of severe trauma or surgery

Narcotic intake within one month of the study and analgesic or muscle relaxants within 24 hours of the study (as self-reported by the examinees).

Hypaesthesia or dysesthesia of the skin – permanent or temporary (as self-reported by the examinees in the general questionnaire and assessed by the examiner during initial visit).

Previous experience of subjects with pressure algometry. (to avoid practice effect bias) [24].

The concept of the study was explained and all subjects completed a consent form prior to participation in the study. All parts of the study were developed within the principles and standards of the Declaration of Helsinki and in accordance with the Guidelines on the Practice of Ethics Committees in Medical Research Involving Human Subjects. The study was designed and approved in accordance with the ethical procedures at the TEI of Lamia, Greece.

Study design – procedures – instruments

A prospective test-retest within-subject repeated measures design was used. In order to assess short-term reliability, subjects were measured on two consecutive days ("Day 1" and "Day 2"). The assessment of the longer-term reliability required another set of

measurements one week after the "Day 1" assessment ("Day 7"). The sPPT measurements were administered by an independent examiner, whereas the initial assessment of the subject was performed by the main investigator (GG). The sPPT measurement involved "pinching" of the skin and subcutaneous tissue without including any muscular tissue.

Protocol: The protocol that was followed for all measurements ("Day 1", "Day 2", "Day 7") is described. Initially, the subjects were given two questionnaires to complete. These two questionnaires (a general one and another to assess psychological distress – see instruments) served to exclude subjects that fulfilled any of the exclusion criteria. The primary investigator recorded blood pressure variables and further examined the subjects for other exclusion criteria (e.g. dysesthesia of the skin). The measuring sites (see later) were then marked on the subjects and the complete procedure was explained. The pressure algometer device was then calibrated according to the procedure described by the manufacturer (by using a pre-determined load and adjusting the algometer accordingly). Practice measurements were then attempted until the subject and the examiner were familiar with the technique. A site different from those included in the main trial was selected for the practice session (e.g. the subject's thenar muscles).

All sPPT measurements were recorded in the same order to prevent order effect bias on the results [25]. Recordings were repeated three times with a five minutes interval between each [26]. Recordings were taken bilaterally with the left side measured first followed by the right side in half of the subjects with the opposite order applied to the rest of the sample, as determined randomly by a coin flip. The mean average of the last two measures was taken as the best estimate of the sPPT. The first measure was discarded since it is shown to usually provide the values with the greatest variance [27].

Standardised instructions were given before each measurement on all occasions. Subjects were instructed to "report as soon as the sensation of pressure changes to pain by pressing the special button attached to the algometer". Usually all procedures were completed within 30–40 min minimising the effect of fatigue [28] and preventing subjects' loss of concentration [29]. Subjects were kept uninformed of their scores throughout the study to prevent previous scores from influencing the results [26]. All measurements took place in the same location (research laboratory), by the same examiner, between 10:00 – 14:00 hrs in order to avoid diurnal variation [30]. The environment (temperature, draughts, cold, dampness) was kept as steady as possible throughout the study in order to control for potential effects on the pain threshold as suggested by Hildebrandt et al. [31]. The examiner remained blind to the previous sPPT values when performing the measurements in order to eliminate the possible effects of examiner's expectancy [32] and to follow the guidelines of successful blinding in reliability studies [33].

The order of sPPT measurements and its topographical location follows:

The T0 point, T2 point and T3 point (Figure 1). The upper trapezius points (T0, T2, T3). Subjects were seated with their arms hanging freely by their sides.

the Ulna point (Figure 2)

the Biceps point (Figure 2)

the Triceps point (Figure 2)

the Gastrocnemius point (Figure 2)

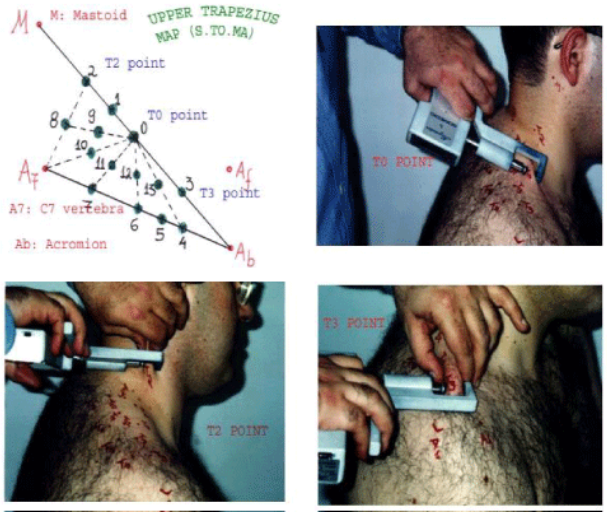

Figure 1: Structured topographical mapping of the upper trapezius area (the S.To.Ma. map)

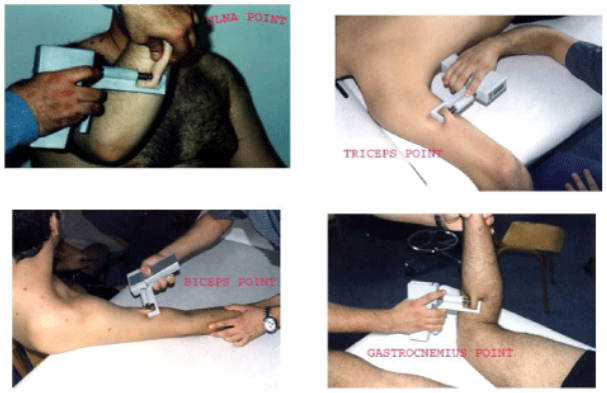

Figure 2: The sPPT measurement sites at Ulna, Biceps, Triceps, Gastrocnemius

Instruments (Device - Questionnaires): The Pressure Algometer: The pressure algometer used in this study was an electronic Somedic type II device (Figure 3). The stimulation unit of the Somedic is gun-shaped with the stimulation tip situated at the end of the barrel, connected to a pressure transducer built into the handle. The measurement units are in kPa/cm2. The accuracy of measurement is 3% and the range used in the current study was from 0-1000 kPa/cm2 (the capacity of the device is up to 2000 kPa/cm2). The device is constructed with a visual display (built into the handle in the form of horizontal flashing light bars), ensuring a smooth and controlled rate of pressure application. The Somedic® algometer is equipped with a push-button that "freezes" the digital display value for 10 sec, when the PPT or sPPT is reached i.e. when the subject pushes the button. This innovation allows for accurate pressure recording of the PPT and sPPT measures [34,35].

The sPPT measurements require that a special attachment is attached to the algometer to allow pinching of the skin and the subcutaneous tissue (Figure 3) [36].

The recording of the sPPT requires that the examiner "pinches" the skin and subcutaneous tissue using this device whilst avoiding any muscular tissue and gradually increasing the pressure with a steady rhythm. When the feeling of pressure changes to discomfort (very early pain – "first pain") the subject presses the button, indicating that the pain threshold has been reached. The absolute value of the measurement depends on several technical characteristics of the algometer and the conditions of the procedure.

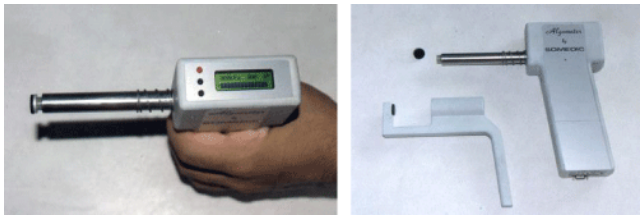

Figure 3: The pressure algometer device used for sPPT measurements

Fischer [37] has suggested measuring sPPT using the same parameters as for measuring PPT. However, from unpublished pilot work, practical experience and feedback from pilot subjects the following parameters were indicated for measuring sPPT:

A constant application rate of approximately 10 kPa/sec in order to allow enough time for the pain sensation to build up steadily and the subject to respond accurately. This rate of application allowed approximately the same response time (3-5 sec) as the typical 50 kPa/sec does for the deeper PPT, which is considered an appropriate time interval.

Selection of the smaller of the available tip-sizes (diameter=0.8 cm). This diameter corresponds to a surface of 0.50 cm^2. List et al. showed that for PPT measurements, the smaller sizes rendered larger PPT measurements possibly due to a spatial summation phenomenon. It is highly likely that this will apply for sPPT measurement as well. Using a smaller surface also permits more precise localisation of the subcutaneous tissue of interest. It would be interesting however, if the manufacturer had provided a smaller surface for the tipsize, to assess Fischer's assertion of a 0.2 cm^2 appropriateness to record skin tenderness [34].

Questionnaires: Two self-reported questionnaires were used in this study: a general and a psychometric one. Both questionnaires were self-reported and administered to the participants prior to the initiation of the study. Each questionnaire took approximately five minutes to complete.

The general questionnaire addressed demographic information and questions regarding the exclusion criteria. The subjects were also asked questions regarding systematic diseases, hypertension and hypotension, and other relevant features. The purpose of this general questionnaire was to identify the appropriate subjects for inclusion in the study.

The psychometric questionnaire: The role of depression and anxiety is relatively clarified in the literature in the determination of PPT and sPPT readings, in the sense that subjects with high levels of psychological distress may respond inappropriately to the examination of PPT and sPPT. In order to identify these individuals, the self-reported Hospital Anxiety and Depression scale (HAD) was administered to all subjects [37]. This questionnaire is considered an adequate alternative to assessing depression when a formal psychological interview is not available [38]. The Greek version of the HAD [39] has evidenced the similar properties of the original version and has been used extensively since its validation (HAD-GR). The HAD-GR consists of 14 questions, seven of which are designed to assess the state of anxiety (HAD-A subscale) while the others provide an insight into the depression levels of the individual (HAD-D subscale). Only the depression sub-scale was analyzed for the purpose of exclusion from the study. Each question according to the answer provides a score from 0-3. Thus, the possible score for depression (7 questions) can range from 0-21. Zigmond and Snaith [37] have suggested a threshold level of eight plus in order to identify potentially pathological cases due to depression. This suggestion has been confirmed by Bjelland et al. [38] in an extensive literature review. In this study a cut-off value of 8 plus was adopted in order to exclude potentially pathological cases. Using this cut-off point, one suspicious case of anxiety and depressive symptomatology was identified and the person was excluded and given advice to seek help to a specialist.

Data analysis

The statistical package SPSS® 10.1 was used for all statistical analyses. The normal distribution of data was assessed using the Shapiro – Wilk test. The Shapiro-Wilk test was selected over the Kolmogorov-Smirnov in order to control for the small sample size (N<50). The descriptive analysis of the variables included the arithmetic mean (mean value), the standard error of the mean (S.E.M.), the standard deviation (S.D.) and the range of measurement (minimum and maximum values).

Intraclass Correlation Co-efficient (ICC) was selected as the appropriate procedure to assess the intra-rater reliability of the sPPT measurements across time (on consecutive days, and after a week) [40]. Specifically, the type ICC (2,1) was used for all intra-rater calculations as suggested by Sim and Wright [40] (p335), since the assessor was the same for all ratings and therefore was considered as a fixed effect (intra-rater reliability) and the mean value of each day measurements was employed in the calculations. Two ICC values were calculated for each measuring site in this study. The first ICC (Day 1 – 2) expresses the reproducibility of measurements for consecutive days (short-term reliability), while the second ICC (Day 1 – 7) describes the longer-term reliability (one week apart). The short-term and longer-term reliability for each measuring site were further visually inspected with scatterplots [41].

Delitto and Strube [42] believe that the ICC's take into account "level" differences, but are not true measures of concordance. Thus, Domholdt [41] suggested that: "…researchers who report reliability on the basis of an ICC should still report the results of an absolute reliability indicator such as the standard error of measurement or the repeated measures ANOVA" (p.287). Thus, an ANOVA model (General Linear Model – Repeated Measures) was further applied to control for potential differences as part of the ICC estimate model.

In order to enhance the rigor of the applied statistic model, two further statistical analyses were performed: the standard error of measurement (SEM) [43] and the smallest detectable difference (SDD) [44]. The SEM derives from the ANOVA error components, is based

on the within-subject error/variability, is a measure of the "precision" of measurement, is expressed in the same units as the original measurement and therefore can be directly compared against subjects' own values [43]. The standard error of measurement (SEM) is a standard deviation of measurement errors and the most frequently calculated statistic for this purpose [41]. It is often estimated as follows:

$$SEM = \sigma\sqrt{1-r}$$

where σ: the standard deviation of the repeated measurements

r: the correlation coefficient (in this study the ICC value)

The SDD is a derivative of the SEM according to the formula:

$$SDD = 1.96\sqrt{2SEM}$$

which can be expressed either in the units of the measurement or as a percentage of the parameter's grand mean. The SDD is a "clinical index" which indicates the level of change in a parameter attributed with 95% certainty to a true change in a subject's condition, instead of being caused by test-retest errors [44]. In a repeated measurements design, the smaller the SDD the more responsive to change a measurement is rendered.

Results

The mean age of the group (N=30) was 25.1 yrs (SD 4.4) [males (N=17): 25.4 yrs (SD 4.5), females (N=13) 24.5 yrs (SD 4.3)].

The distribution of the sPPT values was normal for all variables on most occasions ($p>0.05$). In a couple of instances the statistic reached significance ($p<0.05$) indicating a non-normal distribution, which was not however repeated on subsequent days and was therefore taken as normal.

The descriptive statistics of the sPPT readings are given in Table 1 (mean values, standard deviations, standard error of mean, and the range of values).

Site and Day of Measurement		Mean+Standard Error of Mean (N=30)	Minimum/Maximum	Standard Deviation
	Day 1	53.0 + 4.2	21 - 105	22.9
T0 point	Day 2	52.4 + 3.9	25 - 101	21.4
	Day 7	53.0 +3.8	26 - 98	20.9
	Day 1	47.2 + 3.7	20 - 99	20.4
T2 point	Day 2	47.5 + 3.8	24 - 98	21.1
	Day 7	47.1 + 3.9	19 - 94	21.3
	Day 1	53.2 + 3.7	24 - 106	20.1
T3 point	Day 2	54.8 + 4	23 - 99	22
	Day 7	53.1 + 3.7	21 - 99	20.3
	Day 1	61.1 + 3.8	26 - 101	20.6
Ulna point	Day 2	61.7 + 3.5	32 - 97	19
	Day 7	62.5 + 3.6	34 - 99	19.7
	Day 1	46.0 + 3.8	13 - 103	20.7
Triceps point	Day 2	49.0 + 4	15 - 110	21.6
	Day 7	49.1 + 3.5	16 - 98	19.4
	Day 1	50.4 + 3.6	19 - 84	19.6
Biceps point	Day 2	53.4 + 3.4	22 - 90	18.6
	Day 7	53.9 + 3.4	20 - 89	18.4
	Day 1	51.5 + 3.4	21 - 95	18.4
Gastro point	Day 2	49.7 + 3.9	20 - 103	21.4
	Day 7	50.5 + 3.9	19 - 96	21.3

Table 1: Descriptive statistics of all sPPT data

The ANOVA analysis (repeated measures) did not detect any significant differences among the three days datasets for all sites [$F(1,29) = 0.14 – 2.63$, $p> 0.08$] (Table 2).

sPPT (among the 3 days)	F-value	Signif. p=
T0	0.14	0.87
T2	0.148	0.863
T3	1.701	0.191
Ulna	0.557	0.576
Triceps	1.825	0.17
Biceps	2.629	0.081
Gastrocnemius	0.231	0.794

Table 2: ANOVA repeated-measures used in this case as an absolute reliability indicator. Comparison among the 3 datasets

The ICC (2,1) values for the sPPT showed an excellent repeatability for both short and longer-term reliability. The values ranged for the short-term reliability from 0.75 – 0.97 and for the longer-term from 0.70-0.97 (Tables 3 and 4). Since no differences were shown using ANOVA, a further explorative analysis of the data followed. The Standard Error of Measurement (SEM) and Smallest Detectable Difference (SDD) of each measuring point were calculated. The SEM for the short-term reliability ranged from 3.5 to 10.1 kPa (mean=6.07) and for the longer-term reliability from 3.1 to 11.2 kPa (mean=6.2). The SDD for the short-term reliability ranged from 5.9 to 20 and for the longer-term from 6.5 to 22. This data is summarised in Tables 3 and 4.

Site of Measurement	ICC(2,1) (95% CI) N=30	Grand Mean (kPa)	Std Error of Measurement (kPa)	SDD
T0 point	0.94 (0.87-0.97)	57.7	5.7	10.8
T2 point	0.97 (0.94-0.99)	47.3	3.5	7.3
T3 point	0.97 (0.95-0.99	54	3.2	5.9

Ulna point	0.92 (0.84-0.96)	61.4	5.7	9.3
Triceps point	0.87 (0.75-0.94)	47.5	7.4	15.6
Biceps point	0.86 (0.73-0.93)	51.9	6.9	13.4
Gastro point	0.75 (0.53-0.87)	50.6	10.1	20

Table 3: Short-term reliability measures (Day 1 - Day 2)

Site of Measurement	ICC(2,1) (95% CI) N=30	Grand Mean (kPa)	Std Error of Measurement (kPa)	SDD
T0 point	0.93 (0.88-0.97)	53	5.3	10.1
T2 point	0.97 (0.96-0.99)	47.1	3.1	6.5
T3 point	0.95 (0.90-0.98)	53.2	4.5	8.4
Ulna point	0.94 (0.90-0.98)	61.8	4.6	7.4
Triceps point	0.85 (0.71-0.93)	47.5	7.6	16
Biceps point	0.85 (0.71-0.93)	52.1	7.1	13.6
Gastro point	0.70 (0.44-0.84)	51.1	11.2	22

Table 4: Longer-term reliability measures (Day 1 - Day 7)

The short-term and longer-term reliability of the sPPT readings in the form of scatterplots can be seen in Figure 4. No difference can be observed between short- and longer-term reliability values for the measuring points.

Discussion

The objective of this study was to assess the short and longer-term test-retest reliability of sPPT measurement. The calculated ICC statistic reached good to excellent levels for both types of reliability (0.70–0.97). This is indicative of the property of the measure to reproduce reliable results when the sPPT readings are recorded by the same examiner (intratester-reliability) and over a time interval of up to one week. The clinical significance of the method as assessed by the Smallest Detectable Difference (SDD) statistic was highly acceptable in selected points – trapezius area T0, T2, T3, but less in other – i.e. gastrocnemius.

From the descriptive statistics, it was noted that the sPPT measurements tended to produce similar mean values independent of the area of measurement. This characteristic is indicative of a general measure of skin and subcutaneous tissue sensitivity, which might not be influenced by regional variations and the tissue underneath the measuring site. However, a study measuring sPPT over different tissues (bone, muscle, tendon, ligament, nerve) and at various topographical locations of upper and lower body, without any significant variation, would confirm this finding. Unfortunately, there are no published data until today, to the best of our knowledge, that our subcutaneous pressure threshold measurements can be compared to.

A characteristic of the sPPT measurement, according to this study, is the stability demonstrated over time with the short-term and longer-term reliability demonstrating little variation across time. The high ICC values were cross-examined with a reliability index of concordance – ANOVA (Table 2) and with scatterplot graphs (Figure 4) [41].

The ANOVA revealed no difference among the 3 days measurements and the scatterplots demonstrated pretty consistent plots both for short- and longer-term reliability. Further exploration with the SEM and SDD statistics was necessary in order to understand in real units the actual magnitude of the established reliability and its clinical significance.

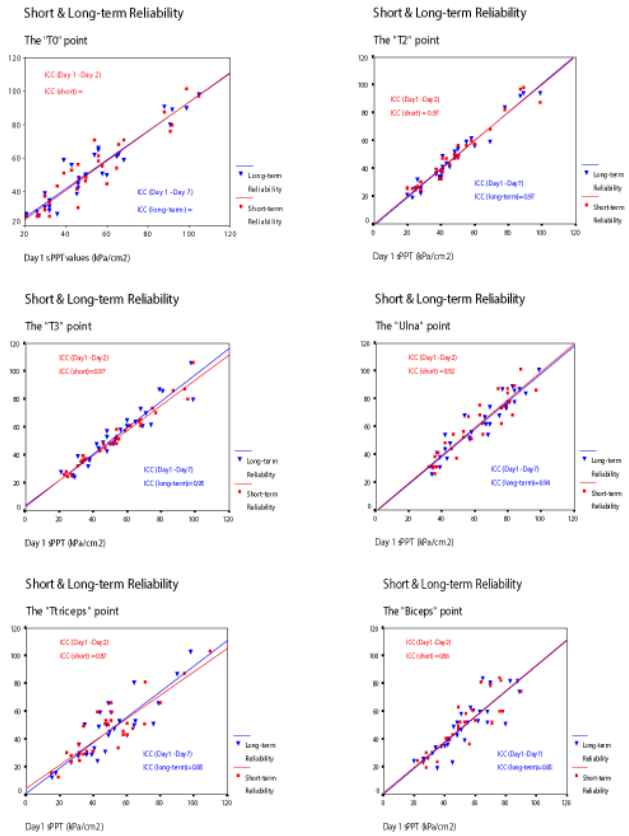

Figure 4: Indicative scatterplots of short and longer-term reliability for the measurement sites.

The sPPT measure produced a relatively high SDD and SEM values in some of the measuring sites. In clinical practice, the SDD's demonstrate the changes required in order for differences to be interpreted as real changes. This could serve for example, in a hypothetical situation, to record the progress of a patient following a treatment approach across time [44]. Therefore, it would be desirable that SDDs are small enough to detect clinically important changes. In this study, the SDDs were relatively high at the non-trapezius sites. However, only healthy volunteers were measured in this attempt and the results cannot be generalised to patients or subjects with suspected changes in skin and subcutaneous tissue mechanical sensitivity. Also, it would be interesting to see if in pathologic states the SDD values differ from the values on normal subjects and further determine the practical clinical usefulness of the measure in patient samples.

In more detail, if someone looks closer to the fluctuation of sPPT data in this study compared with the literature on PPT (deep PPT) in

general, realizes that despite the fact that SDD values may be high, sPPT measurements still are better (lower) than the usual PPT SDD values. Specifically, for the trapezius area (T0, T2, T3) the sPPT SDD ranged from 5.9% to 10.8% for the short-term and from 6.5% to 10.1% for the longer-term reliability. This variability is significantly less than that described for PPT measures where a 18.8% to 28.5% difference in the mean average PPT value is a typical example of pressure algometry measurements [45]. Therefore, in pragmatic terms the SDD values for the sPPT variation may be considered acceptable, especially for clinical studies and when compared to usual PPT SDD values.

The trapezius sPPT measures produced higher ICC and smaller SDD values than the other measuring sites. This could be due to the easier accessibility of the trapezius points (T0, T2, T3) compared for example, to the gastrocnemius or biceps points. The T0, T2, T3 and Ulna sPPT points posed no problems to the examiner when locating them and distinguishing the skin and subcutaneous tissues from the muscle or bone underneath. Sometimes, this was practically more difficult to determine for the gastrocnemius, biceps and triceps points. Another important factor that may have influenced the specific sites differences was the localisation procedure. For example, the localisation of T0, T2 and T3 points followed a highly standardised procedure. A complete mapping of the area was first drawn and thereafter the T0, T2 and T3 points were marked. In this way, the points were located in relation to the whole trapezius area and not only in relation to a reference site. On the other hand, all other points were marked in reference to nearby bony landmarks. Although it cannot be practically demonstrated, this procedure may have produced the consistently higher ICC and lower SDD values for the trapezius area (T0, T2, T3). Possibly, a similarly highly standardised approach for the rest of measuring points would produce ICC and SDD values comparable to the trapezius area.

Nevertheless, a number of additional factors may have contributed to the good reliability statistics, such as:

The experience of the examiner with pressure algometry and furthermore with subcutaneous PPT [46].

The technical parameters adopted for the sPPT measurement (tipsize, rate of pressure application) derived from the clinical experience of the examiner and existing literature

The methodology of the study that controlled for a great range of potentially influential factors (i.e. diurnal variations etc), aiming to take into consideration all known influential factors for pressure algometry measurements

In summary, these results imply that subcutaneous pressure algometry can produce reliable readings across time, when the same examiner performs the measurements under the steady conditions that were followed in this study. A highly standardised approach seems to be necessary, if clinically meaningful measurements with reduced error variance are desired.

Future studies are needed to assess if reliability of sPPT remains similarly high after longer periods of time (e.g. 2 weeks or 1 month). The inter-observer sPPT reliability needs also to be demonstrated so that subcutaneous algometry findings have the potential to be interchangeable among examiners in a future clinical setting. Additionally, it remains within the scope of future studies to expand the present findings in pathologic states and to establish the clinical usefulness (SDD statistic) in those samples, as well.

Conclusions

The aim of this study was to discuss the role of subcutaneous pressure pain threshold measurements in musculoskeletal assessment by updating its current state, but mainly to investigate its test-retest reliability (intra-tester) in the short and longer-term periods. As it was shown by the results of this study, subcutaneous pressure pain threshold measurements (sPPT) seem to be reliable and reproducible measurements, in the short-term (between days) and in the longer term (after a week) as well. The sPPT statistical properties as described herein are similar to the usual deep PPT measurements. It remains to implement the sPPT measurements in clinical conditions in order to clarify its potential role as a treatment outcome measure and an index of patient progression.

References

1. Adnadjevic D, Graven-Nielsen T (2014) Vibration and Rotation During Biaxial Pressure Algometry Is Related with Decreased and Increased Pain Sensations. Pain Med.
2. Falk S, Ipsen DH, Appel CK, Ugarak A, Durup D, et al. (2014) Randall Selitto pressure algometry for assessment of bone-related pain in rats. Eur J Pain.
3. Reid KI, Carlson C, Rayens MK, Gracely RH (1996) The influence of cutaneous tissue afferents on masticatory pain-pressure thresholds. J Orofac Pain 10: 324-329.
4. Sessle BJ, Hu JW, Amano N, Zhong G (1986) Convergence of cutaneous, tooth pulp, visceral, neck and muscle afferents onto nociceptive and non-nociceptive neurones in trigeminal subnucleus caudalis (medullary dorsal horn) and its implications for referred pain. Pain 27: 219-235.
5. Jaeger B, Reeves JL (1986) Quantification of changes in myofascial trigger point sensitivity with the pressure algometer following passive stretch. Pain 27: 203-210.
6. Fujisawa M, Shoji S, Ishibashi K, Clark GT (1999) Pressure pain threshold with and without iontophoretic anesthesia of the masseter muscle in asymptomatic males. J Orofac Pain 13: 97-103.
7. Sarrafzadeh J, Ahmadi A, Yassin M (2012) The effects of pressure release, phonophoresis of hydrocortisone, and ultrasound on upper trapezius latent myofascial trigger point. Arch Phys Med Rehabil 93: 72-77.
8. Valenza MC, Torres-Sanchez I, Cabrera- Martos I, Valenza-Demet G, Cano-Cappellacci M (2013) Acute Effects of contract-relax stretching vs. TENS in young subjects with anterior knee pain: a randomized controlled trial. J Strength Cond Res, Nov 20.
9. Affaitati G, Fabrizio A, Savini A, Lerza R, Tafuri E, et al. (2009) A randomized, controlled study comparing a lidocaine patch, a placebo patch, and anesthetic injection for treatment of trigger points in patients with myofascial pain syndrome: evaluation of pain and somatic pain thresholds. Clin Ther 31: 705-720.
10. Baldry P (1995) Superficial dry needling at myofascial trigger point sites. Journal of Musculoskeletal Pain 3: 117-126.
11. Jensen K, Andersen HO, Olesen J, Lindblom U (1986) Pressure-pain threshold in human temporal region. Evaluation of a new pressure algometer. Pain 25: 313-323.
12. Kosek E, Ekholm J (1995) Modulation of pressure pain thresholds during and following isometric contraction. Pain 61: 481-486.
13. Vecchiet L, Galletti R, Giamberardino MA, Dragani L, Marini F (1988) Modifications of cutaneous, subcutaneous, and muscular sensory and pain thresholds after the induction of an experimental algogenic focus in the skeletal muscle. Clinical Journal of Pain 4: 55-59.
14. Giamberardino MA, Dragani L, Vecchiet L (1988) Measurement of sensory and pain thresholds in the subcutaneous tissue by means of electrical stimulation. Pain Clinic 2: 41-44.
15. Vecchiet L, Giamberardino MA, Dragani L, De Bigontina P, Fessard AD, et al. (1990) Latent myofascial trigger points: Changes in muscular and

subcutaneous pain thresholds at trigger point and target level. Journal of Manual Medicine 5: 151-154.

16. Vecchiet L, Giamberardino MA (1997) Referred pain: clinical significance, pathophysiology and treatment. Physical Medicine and Rehabilitation Clinics of North America 8: 119-136.
17. Fischer AA (1998) Algometry in diagnosis of musculoskeletal pain and evaluation of treatment outcome: An update. Journal of Musculoskeletal Pain 6: 5-32.
18. Rachlin ES (1994) Myofascial pain and fibromyalgia, 1st edn. St. Louis: Mosby.
19. Vecchiet L, Giamberardino MA, Saggini R (1991) Myofascial pain syndromes: clinical and pathophysiological aspects. Clinical Journal of Pain 7: S16-S22
20. List T, Helkimo M, Karlsson R (1993) Pressure pain thresholds in patients with craniomandibular disorders before and after treatment with acupuncture and occlusal splint therapy: a controlled clinical study. Journal of Orofacial Pain 75: 275-282.
21. Agostino R, Cruccu G, Iannetti G, Romaniello A, Truini A, et al. (2000) Topographical distribution of pinprick and warmth thresholds to CO2 laser stimulation on the human skin. Neurosci Lett 285: 115-118.
22. Walter SD, Eliasziw M, Donner A (1998) Sample size and optimal designs for reliability studies. Stat Med 17: 101-110.
23. Song JY, Merskey H, Noh S, Sullivan S (1993) The effect of controlling for anxiety and depression upon the threshold for pressure pain in three comparison groups. Journal of Musculoskeletal Pain 1(1): 73.
24. Hicks CM (1995) Research for physiotherapists. Project design and analysis., 2nd edn. Edinburg: Churchill Livingstone.
25. Puttick M, Schulzer M, Kinkhoff A, Koehler B, Rangno K, et al. (1995) Reliability and reproducubility of fibromyalgic tenderness, measurement by electronic and mechanical dolorimeters. Journal of Musculoskeletal Pain 3: 3-14.
26. Delaney GA, McKee AC (1993) Inter- and intra-rater reliability of the pressure threshold meter in measurement of myofascial trigger point sensitivity. Am J Phys Med Rehabil 72: 136-139.
27. Pontinen PJ (1998) Reliability, validity, reproducibility of algometry in diagnosis of active and latent tender spots and trigger points. Journal of Musculoskeletal Pain 6: 61-71.
28. Hagberg M, Kvarnström S (1984) Muscular endurance and electromyographic fatigue in myofascial shoulder pain. Arch Phys Med Rehabil 65: 522-525.
29. Hagander LG, Midani HA, Kuskowski MA, Parry GJ (2000) Quantitative sensory testing: effect of site and skin temperature on thermal thresholds. Clin Neurophysiol 111: 17-22.
30. Koltyn KF, Focht BC, Ancker JM, Pasley J (1999) Experimentally induced pain perception in men and women in the morning and evening. Int J Neurosci 98: 1-11.
31. Hildebrandt VH, Bongers PM, van Dijk FJ, Kemper HC, Dul J (2002) The influence of climatic factors on non-specific back and neck-shoulder disease. Ergonomics 45: 32-48.
32. Ohrbach R, Crow H, Kamer A (1998) Examiner expectancy effects in the measurement of pressure pain thresholds. Pain 74: 163-170.
33. Day SJ, Altman DG (2000) Statistics notes: blinding in clinical trials and other studies. BMJ 321: 504.
34. Fischer AA (1986) Pressure threshold meter: its use for quantification of tender spots. Arch Phys Med Rehabil 67: 836-838.
35. Fischer AA (1986) Pressure tolerance over muscles and bones in normal subjects. Arch Phys Med Rehabil 67: 406-409.
36. Fischer AA (1997) New developments in diagnosis of myofascial pain and fibromyalgia. Physical Medicine and Rehabilitation Clinics of North America 8: 1-22.
37. Zigmond AS, Snaith RP (1983) The hospital anxiety and depression scale. Acta Psychiatr Scand 67: 361-370.
38. Bjelland I, Dahl AA, Haug TT, Neckelmann D (2002) The validity of the Hospital Anxiety and Depression Scale. An updated literature review. J Psychosom Res 52: 69-77.
39. Georgoudis G, Oldham JA (2001) Anxiety and Depression as confounding factors in cross-cultural pain research studies: Validity and reliability of a Greek version of the Hospital Anxiety & Depression Scale. Physiotherapy 87: 92-93.
40. Sim J, Wright C (2000) Research in health care: concepts, designs and methods. Stanley Thornes Ltd.
41. Domholdt E (1993) Physical Therapy Research, 1st edn. Philadelphia: W.B. Saunders Company.
42. Delitto A, Strube MJ (1991) Reliability in the clinical setting. Research Section Newsletter 24: 2-8.
43. Rankin G, Stokes M (1998) Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. Clin Rehabil 12: 187-199.
44. Roebroeck ME, Harlaar J, Lankhorst GJ (1993) The application of generalizability theory to reliability assessment: an illustration using isometric force measurements. Phys Ther 73: 386-395.
45. Nussbaum EL, Downes L (1998) Reliability of clinical pressure-pain algometric measurements obtained on consecutive days. Phys Ther 78: 160-169.
46. Chung SC, Um BY, Kim HS (1992) Evaluation of pressure pain threshold in head and neck muscles by electronic algometer: intrarater and interrater reliability. Cranio 10: 28-34.