

ORIGINAL REPORT

## MEASUREMENT VARIABILITY OF QUANTITATIVE SENSORY TESTING IN PERSONS WITH POST-STROKE SHOULDER PAIN

Ingrid Lindgren, RPT, PhD<sup>1,2</sup>, Elisabeth Ekstrand, RPT, MSc<sup>2,3</sup> and Christina Brogårdh, RPT, PhD<sup>2,1</sup>

From the <sup>1</sup>Department of Neurology and Rehabilitation Medicine, Skåne University Hospital, <sup>2</sup>Department of Health Sciences, Physiotherapy, Lund University, Lund and <sup>3</sup>Department of Hand Surgery, Skåne University Hospital, Malmö, Sweden

**Objective:** To evaluate the measurement variability of quantitative sensory testing (QST) in persons with post-stroke shoulder pain.

**Design:** A test-retest design.

**Participants:** Twenty-three persons with post-stroke shoulder pain (median age 65 years).

**Methods:** Thermal detection thresholds (cold and warm), pain thresholds (cold and heat) and mechanical pain thresholds (pressure and pin prick) were assessed twice in both arms, 2–3 weeks apart. Measurement variability was analysed with the intraclass correlation coefficient (ICC<sub>2,1</sub>), the change in mean ( $\bar{d}$ ) with 95% confidence interval (logarithmic scales), and the relative standard error of measurement (SEM%; re-transformed scales).

**Results:** The ICCs for thermal thresholds ranged from 0.48 to 0.89 in the affected (painful) arm and from 0.50 to 0.63 in the unaffected arm, and for mechanical pain thresholds from 0.66 to 0.90 in both arms. No systematic changes in the mean ( $\bar{d}$ ) were found. The SEM% ranged from 4% to 10% for thermal detection and heat pain thresholds, and from 17% to 42% for cold pain and mechanical pain thresholds in both arms.

**Conclusion:** QST measurements, especially cold pain thresholds and mechanical pain thresholds, vary in persons with post-stroke shoulder pain. Before QST can be used routinely to evaluate post-stroke shoulder pain, a test protocol with decreased variability needs to be developed.

**Key words:** shoulder pain; stroke; sensory thresholds; intra observer variability.

J Rehabil Med 2016; 48: 435–441

*Correspondence address:* Ingrid Lindgren, Department of Health Sciences, Physiotherapy, Lund University, Box 157, SE-221 00 Lund, Sweden. E-mail: [ingrid.lindgren@med.lu.se](mailto:ingrid.lindgren@med.lu.se)

Accepted Mar 8, 2016; Epub ahead of print Apr 15, 2016

### INTRODUCTION

Post-stroke shoulder pain (PSSP) is one of the most common types of pain after stroke (1, 2). It occurs in approximately one-third of an unselected stroke population and in half of persons with persistent upper extremity paresis (3). PSSP is often a long-lasting problem (4) and effective treatments are lacking.

PSSP has mostly been associated with reduced motor function (3, 5–7) and decreased range of motion in the shoulder (4, 8). Some studies indicate that somatosensory impairments, such as pain hypersensitivity, thermal and mechanical hyperalgesia and allodynia, may contribute to PSSP, which could indicate the presence of a neuropathic component (9–11). However, in our recently published study using quantitative sensory testing (QST), no differences were found in somatosensory impairments between persons with and without PSSP (12), implying that PSSP is not associated with neuropathic mechanisms. Plausible explanations for the different results could be differences in study designs; for example, measurement locations, pain characteristics among the participants and the choice of measurements variables.

QST is an established method to assess somatosensory modalities, such as thermal detection and pain thresholds as well as mechanical pain thresholds (13). QST has been used in different populations (14–17), including studies of PSSP (9–12). Thermal thresholds have been found to be reliable in persons with diabetes (18), complex regional pain syndrome (CRPS) (19), spinal cord injury (16), Alzheimer's disease (15) and in healthy individuals (20–22). Mechanical thresholds are shown to be reliable in persons with Alzheimer's disease (15) as well as in healthy individuals (23, 24). To the best of our knowledge, no study has investigated the reliability (measurement variability) of QST measurements in persons with PSSP. This knowledge is important to facilitate the interpretation of changes in somatosensory impairments in PSSP over time or after an intervention.

The purpose of this study was therefore to evaluate the measurement variability of QST for thermal detection and thermal pain thresholds as well as for mechanical pain thresholds in persons with PSSP.

### METHODS

#### *Participants*

A total of 23 persons with PSSP participated in this study. They were recruited from the Lund Stroke Register and the Department of Neurology and Rehabilitation Medicine at Skåne University Hospital by screening medical records. Inclusion criteria were: stroke onset between 5 and 36 months prior to study enrollment, decreased motor function in the affected arm and daily pain in the affected shoulder for

a duration of at least 4 months after stroke onset. Exclusion criteria were: difficulty in communicating or in understanding test instructions; other conditions that caused pain or sensory disturbances; and severe depression.

At the time of the assessment, the participants were included in another study with the purpose of investigating whether somatosensory impairments were more common in individuals with PSSP in comparison with individuals without PSSP and healthy controls (12). The present study included only those with PSSP.

#### Ethics

Prior to inclusion, information about the purpose of the study was provided and each person provided written consent to participate. The principles of the Declaration of Helsinki were followed, and the study was approved by the Regional Ethical Review Board, Lund, Sweden (Dnr 2011/742).

#### Demographics and participant characteristics

Before the QST measurements, age, sex, body mass index (BMI), prescribed antidepressant medication and pain medication were recorded. Pain medication was unchanged at the time of assessment. Stroke-specific characteristics, such as side of lesion, type of stroke, stroke onset, independence in personal activities of daily living (P-ADL), shoulder pain intensity, upper extremity motor function, light touch, proprioception and abnormal somatosensation in the affected side, were registered.

Shoulder pain intensity during active movements was registered with the Visual Analogue Scale for Pain (VAS-P) (25) and motor function of the upper extremity with the Swedish version of the Modified Motor Assessment Scale (M-MAS) (26, 27). Light touch in the upper arms and forearms, hands and fingers, as well as proprioception in the thumbs (interphalangeal joints) and wrists were assessed with the Fugl-Meyer Assessment of Sensorimotor Recovery After Stroke (28). These outcome measurements are described in detail in our previous study (12).

#### Procedure

All QST measurements were performed on 2 test occasions, 2–3 weeks apart, by one of the authors (EE). Each test session was performed in a quiet hospital room and lasted approximately 1 h. During assessments, the participants were seated in a chair with armrests. The upper arm was placed in a comfortable position with the elbow at approximately 90° flexion and the forearm resting on a pillow. Participants were not allowed to view the computer screen during the tests. Prior to the study, the QST protocol was standardized with regard to order of the assessments and number of repetitions, and was carefully adhered to the patient group in order to avoid fatigue. On the second test occasion, the examiner was unable to see the QST measurements from the first test occasion.

#### Quantitative sensory testing

The thermal thresholds were measured with the MSA Thermostest (Somedic AB, Hörby, Sweden, <http://www.somedic.com/>). The 4 thermal tests were: cold detection thresholds (CDT); warm detection thresholds (WDT); cold pain thresholds (CPT) and heat pain thresholds (HPT). The 2 mechanical tests included: pressure pain thresholds (PPT) and pin-prick pain thresholds (PPPT), measured by the Algometer and the SenseBox Electronic von Frey (Somedic AB, Hörby, Sweden), respectively. First, the unaffected arm was measured and thereafter the affected (painful) arm. During assessments the intensity of the stimulus applied to the skin was increased/decreased until the detection threshold (minimum intensity of a stimulus perceived as stimulus) or pain threshold (minimum intensity of a stimulus perceived as painful) was reached (i.e. “the method of limits”). The participant held a switch in the unaffected hand and when the detection threshold (CDT, WDT) or pain threshold (CPT, HPT, PPT, PPPT) was perceived, he or she pressed the switch, at which point the assessment was stopped and the threshold value digitally recorded.

**Thermal testing.** The thermal thresholds were performed in the following order: cold detection, warm detection, cold pain and heat pain. A thermode, 25 × 50 mm, with an initial temperature of 32°C and a rate of change of temperature of 1°/s, was applied to the skin on the upper arm. Measurements were made with a precision of 0.1°. During the cold tests the temperature gradually decreased, to a minimum of 10°C. During the warmth/heat tests the temperature gradually increased, to a maximum of 50°C. A higher CDT or CPT indicated that cold or cold pain was perceived at a lower temperature and a higher WDT or HPT that warmth or heat pain was perceived at a higher temperature.

To become familiar with the test, the thermal test was performed on a control point on the unaffected leg over the m. vastus medialis (distal part). Thereafter, the measurements were performed on the upper arm over the middle part of the middle portion of the deltoid muscle. In all locations, 4 repetitions were performed with a 4–6 s interval between each repetition. Results are presented as medians of the 4 assessments for each variable.

**Mechanical testing.** The PPT was measured with an electronic algometer. A probe with a pressure diameter of 1 cm<sup>2</sup>, a slope of 50 kPa/s, and an initial pressure of 10 kPa was applied to 3 points on the upper arm: upper, middle and lower part of the mid-deltoid muscle. The measurements were made with a precision of 0.1 kPa. The pressure was gradually increased until the subject indicated pain perception. Maximum pressure that could be applied was preset to 1,000 kPa. Two PPT assessments were performed at each of the 3 points, yielding a total of 6 measurements. The results are presented as the median of the 6 measurements.

The PPPT was measured with an electric von Frey transducer, using a 0.2-mm tip diameter, a speed of 10 g/s, with an initial pressure of 10 g at 3 points on both the anterior and posterior part of the deltoid muscle. The measurements were made with a precision of 0.1 g. The pressure was gradually increased until the subject indicated pain and the maximum pressure that could be applied was 400 g. A higher PPPT indicated that pin-prick pain was perceived at a greater pressure. The PPPT was measured once, yielding 6 measurements for each arm. The results are presented as the median of the 6 measurements.

#### Statistical methods

Demographic data and characteristics for the PSSP group are presented as frequencies and medians (minimum–maximum), and the QST measurements as medians (minimum–maximum).

As the differences between repeated OST measurements had a heteroscedastic pattern when graphically analysed, log-transformations ( $\log_{10}$ ) were performed. To determine the measurement variability, several statistical methods were applied (29, 30): the intraclass correlation ( $ICC_{2,1}$ ); the change in the mean ( $d$ ); Bland & Altman graphs and the relative standard error of measurement (SEM%).

The  $ICC_{2,1}$  was calculated on the logarithmic scale, together with 95% confidence interval (CI) (31). The change in the mean ( $d$ ), with a 95% CI (logarithmic scale) between the 2 measurements (test 2 minus test 1) was calculated to determine whether there was a true systematic difference. If zero was included in the 95% CI for  $d$ , it was not indicative of a systematic change (32). To visually interpret the data, Bland & Altman graphs were formed (on the logarithmic scale). The difference from the 2 test occasions was plotted against the mean of the 2 test occasions for each participant and presented as  $d$  with 95% limits of agreement (LOA,  $d \pm 1.96 \times$  standard deviation). The SEM gives the measurement variability in absolute values and indicates the extent of the measurement error caused by random variation for a group of individuals. The SEM was calculated as the square root of the total within-subject error variance. From the SEM, the SEM% was calculated, representing the standard error of measurement in relative terms, i.e. the percentage of the mean of all value of SEM-1 (i.e.  $10^{SEM} - 1$ ). Thus, the SEM% was presented as original values after re-transformation from logarithmic values.

Data were analysed using the IBM SPSS Statistics version 22 (IBM Corp., Armonk, New York, USA).

Table I. Demographics and characteristics of the 23 persons with post-stroke shoulder pain

Demographics and characteristics	
Age, years, median (min–max)	65 (45–81)
Male, <i>n</i> (%)	19 (83)
BMI, median (min–max)	28 (22–35)
Right-hemispheric lesion, <i>n</i> (%)	11 (48)
Stroke type, <i>n</i> (%)	
Cerebral infarction	19 (83)
Haemorrhage	4 (17)
Stroke onset, months, median (min–max)	13 (5–33)
Independency in P-ADL, <i>n</i> (%)	22 (96)
Shoulder pain intensity in active movements <sup>a</sup> , mm, median (range)	46 (10–100)
Upper extremity motor function in affected (painful) side <sup>b</sup> , <i>n</i> (%)	
0–11 points, severe–moderate restriction	18 (78)
12–15 points, no–mild restriction	5 (22)
Light touch abnormal or absent in affected (painful) arm <sup>c</sup> , <i>n</i> (%)	6 (26)
Proprioception decreased or absent in affected (painful) arm <sup>d</sup> , <i>n</i> (%)	5 (22)
Abnormal somatosensation in affected (painful) side; <i>n</i> (%)	11 (48)
Prescribed SSRI, <i>n</i> (%)	4 (17)
Prescribed pain medication, <i>n</i> (%)	3 (13)

<sup>a</sup>Visual Analogue Scale for Pain, VAS-P, score 0–100 mm; <sup>b</sup>Modified Motor Assessment Scale, assessed in the upper arm and hand and as advanced hand activities, score 0–15 points; <sup>c</sup>According to Fugl-Meyer, assessed in the upper arm and forearm, hands and fingers; <sup>d</sup>According to Fugl-Meyer, assessed in the thumbs and wrists. Min: minimum; max: maximum. SSRI: selective serotonin re-uptake inhibitors.

RESULTS

Demographics and participant characteristics

Baseline demographics and characteristics are presented in Table I. A majority (78%) of the participants had moderate-to-severe restrictions in upper extremity motor function and 48% reported abnormal somatosensation in the affected side. Median shoulder pain intensity was 46 mm (minimum–maximum: 10–100 mm). Four participants were taking prescribed selective serotonin re-uptake inhibitors (SSRI) and 3 were taking prescribed pain medications, such as paracetamol and opioids.

The median values for all QST measurements (thermal thresholds and mechanical thresholds) on both test occasions are presented for the affected (painful) arm and the unaffected arm, respectively in Table II. The median interval between test occasion 1 and 2 was 18 days (minimum 13 days, maximum 36 days).

Quantitative sensory testing

The measurement variabilities for all QST measurements for both the affected (painful) arm and the unaffected arm are shown in Table III.

Thermal thresholds

For the thermal thresholds, the ICCs ranged from 0.48 to 0.89 in the affected (painful) arm and from 0.50 to 0.63

Table II. Quantitative sensory testing of the affected (painful) and unaffected arm for the 23 persons with post-stroke shoulder pain

	Test occasion 1 Median (min–max)	Test occasion 2 Median (min–max)
CDT, °C		
Affected (painful) arm	28.8 (10.0–30.3)	29.1 (10.0–31.0)
Unaffected arm	29.8 (12.7–31.3)	30.0 (25.1–31.2)
WDT, °C		
Affected (painful) arm	38.6 (34.5–50.0)	38.1 (24.4–50.0)
Unaffected arm	39.5 (34.7–49.5)	39.5 (24.7–49.5)
CPT, °C		
Affected (painful) arm	10.0 (10.0–28.5)	10.0 (10.0–28.5)
Unaffected arm	10.0 (10.0–27.3)	12.3 (10.0–25.9)
HPT, °C		
Affected (painful) arm	48.8 (43.4–50.0)	48.8 (43.4–50.0)
Unaffected arm	48.1 (39.6–50.0)	48.2 (39.6–50.0)
PPT, kPa		
Affected (painful) arm	359 (182–1000)	434 (186–1000)
Unaffected arm	370 (205–1000)	422 (167–1000)
PPPT, g		
Affected (painful) arm	220 (59–400)	189 (45–400)
Unaffected arm	151 (26–400)	162 (44–400)

CDT: cold detection threshold; WDT: warm detection threshold; CPT: cold pain threshold; HPT: heat pain threshold; PPT: pressure pain threshold; PPPT: pin-prick pain threshold; Min: minimum; max: maximum.

in the unaffected arm. For both arms, all *d* values were generally low, the widths of the 95% CI for *d* were narrow and no systematic changes in the mean (*d*) were found. The SEM% in both the affected (painful) arm and the unaffected arm ranged from 4% to 10% for CDT, WDT and HPT, but was 33% for CPT in both arms.

Table III. Measurement variability presented as intraclass correlation coefficient (ICC2.1), the mean difference (*d*) together with the 95% confidence interval (CI) for *d*, and the relative standard error of measurement (SEM%) for the affected (painful) and unaffected arm in the 23 persons with post-stroke shoulder pain

	ICC (95% CI)	<i>d</i> (95% CI)	SEM %
CDT			
Affected (painful) arm	0.89 (0.75–0.95)	–0.005 (–0.031–0.022)	10
Unaffected arm	0.51 (0.14–0.76)	0.009 (–0.015–0.034)	10
WDT			
Affected (painful) arm	0.56 (0.20–0.79)	0.001 (–0.019–0.020)	7
Unaffected arm	0.58 (0.25–0.80)	–0.009 (–0.022–0.004)	5
CPT			
Affected (painful) arm	0.48 (0.09–0.74)	0.018 (–0.061–0.096)	33
Unaffected arm	0.50 (0.11–0.75)	–0.003 (–0.080–0.075)	33
HPT			
Affected (painful) arm	0.52 (0.16–0.76)	–0.005 (–0.015–0.004)	4
Unaffected arm	0.63 (0.31–0.82)	–0.004 (–0.013–0.006)	4
PPT			
Affected (painful) arm	0.90 (0.77–0.96)	–0.002 (–0.045–0.040)	17
Unaffected arm	0.89 (0.76–0.95)	–0.002 (–0.045–0.040)	17
PPPT			
Affected (painful) arm	0.66 (0.34–0.84)	–0.003 (–0.098–0.091)	42
Unaffected arm	0.77 (0.52–0.89)	0.001 (–0.082–0.084)	36

CDT: cold detection threshold; WDT: warm detection threshold; CPT: cold pain threshold; HPT: heat pain threshold; PPT: pressure pain threshold; PPPT: pin-prick pain threshold. ICC and *d* are calculated on logarithmic values (log10). SEM% is presented as original values after re-transformation from logarithmic values.

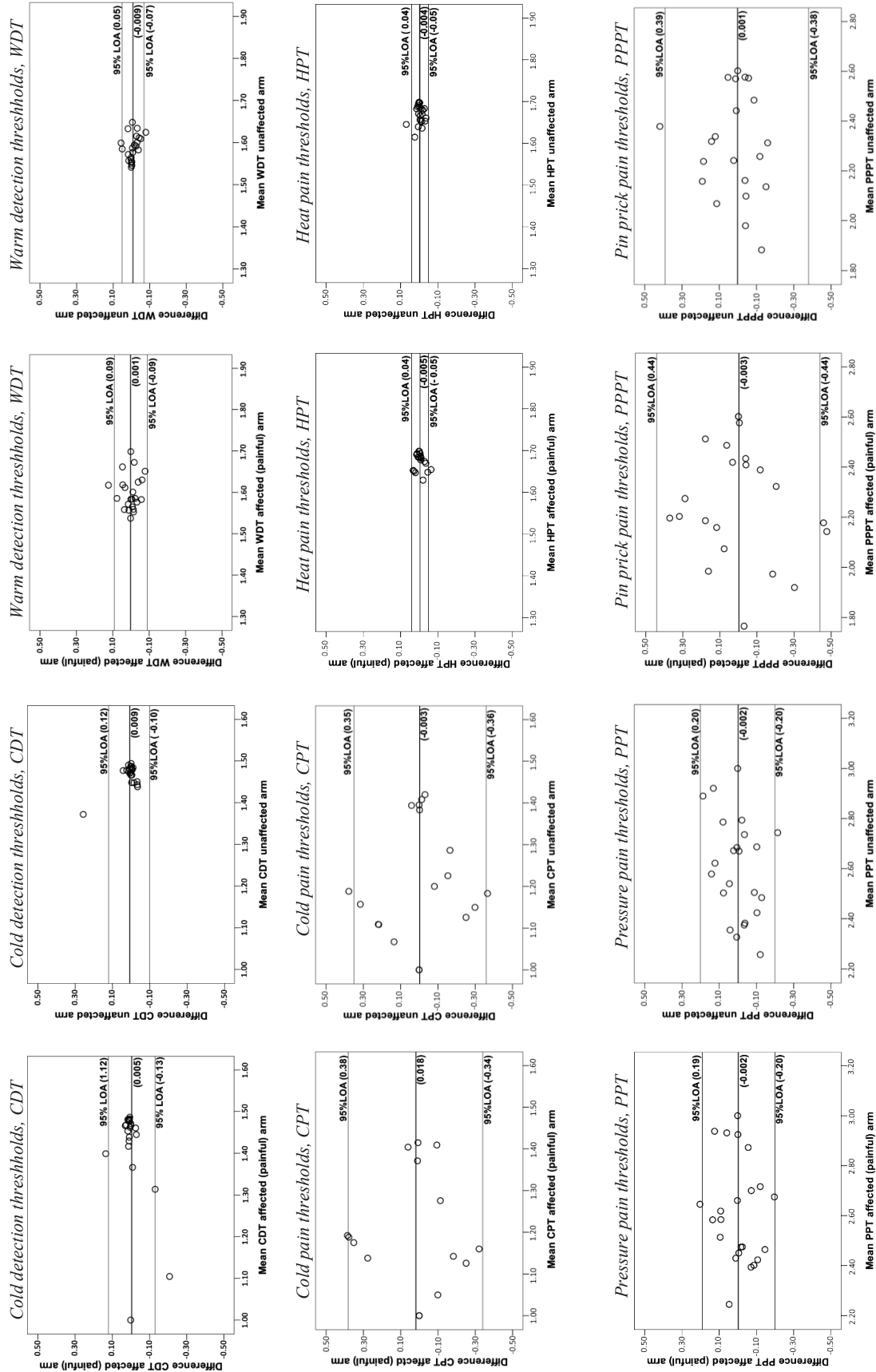


Fig. 1. Bland & Altman graphs (logarithmic data) illustrate the differences between the 2 test occasions plotted against the mean of the 2 test occasions for the affected (painful) and unaffected arm. The changes in the mean (d) (bold lines) and the 95% limits of agreement (LOA) are illustrated in the graphs for thermal detection thresholds (cold detection thresholds (CDT) and warm detection thresholds (WDT)), pain thresholds (cold pain thresholds (CPT) and heat pain thresholds (HPT)) as well as mechanical pain thresholds (pressure pain thresholds (PPT) and pin-prick pain thresholds (PPPT)).

### Mechanical thresholds

For the mechanical thresholds, the ICCs ranged from 0.66 to 0.90 in the affected (painful arm) and from 0.77 to 0.89 in the unaffected arm. For both arms, the *d* values were generally low and the widths of the 95% CI for *d* were narrow. No systematic changes in the mean (*d*) were found for the mechanical thresholds. The SEM% for PPT was 17% for both arms. For PPPT, the SEM% was 42% for the affected (painful) arm and 36% for the unaffected arm, respectively.

In the Bland & Altman graphs (Fig. 1), the changes in the mean (*d*) and the 95% LOA are presented. For both the affected (painful) arm and the unaffected arm, the 95% LOA ranged approximately from -0.1 to 0.1 for CDT, WDT and HPT; from -0.2 to 0.2 for PPT and from -0.3 to 0.4 for CPT and PPPT.

## DISCUSSION

To the best of our knowledge, this is the first study that has assessed the measurement variability of QST measurements for thermal detection and pain thresholds, as well as for mechanical pain thresholds in persons with PSSP. The ICCs for both the affected (painful) and unaffected arm ranged from 0.48 to 0.90, with no systematic changes in the mean. The SEM% was lowest for CDT, WDT and HPT (4–10%), but higher for CPT, PPT and PPPT (17–42%).

Overall, the ICCs of QST measurements (logarithmic scale) varied considerably in our study. Except for a difference in ICC for CDT (0.89 for the affected (painful) arm and 0.51 for the unaffected arm), only minor differences were seen between the arms. Previous test-retest reliability studies of QST measurements have used different statistical methods and variables on both the original and the logarithmic scales, and therefore our results are difficult to fully compare with those studies. However, Jensen-Dahm et al. (15) evaluated the reliability of upper extremity thermal and mechanical thresholds (log transformation of some variables) in persons with mild-to-moderate Alzheimer's disease and in healthy controls. In the Alzheimer group, the ICCs ranged from 0.32 to 0.50 and in the control group from 0.50 to 0.84. Another study investigated the reliability of QST measurements in persons with spinal cord injury and healthy controls (16). They found that ICCs for CDT and WDT (logarithmic values) ranged from 0.90 to 0.95 in the spinal cord group and from 0.68 to 0.70 in the healthy controls, while CPT and HPT (original values) was 0.50 in the spinal cord group and ranged from 0.49 to 0.68 in the healthy controls. Taken together, this indicates that ICCs of repeated QST measurements vary considerably in different study populations.

In the present study, a majority of the ICCs for thermal thresholds were rather low (0.48–0.63 except for CDT in the affected (painful) arm which was 0.89) whereas the mechanical thresholds were higher (0.66–0.90). A possible reason for the low ICCs in thermal thresholds could be the measurement distribution of the participants. As the ICC is dependent on the between subject variability, a narrow distribution of measurements (i.e. a low between subject variability) can yield a false

low ICC even if the agreement is high (30, 31). As can be seen in Table III and Fig. 1, narrow distributions of measurements were seen in both arms in WDT and HPT and, consequently, the ICCs were low. Moreover, in CDT the measurement distribution was narrower in the unaffected arm compared with the affected (painful) arm, which might explain the lower ICC for the unaffected arm. Conversely, a wide distribution of measurements (i.e. a high between-subject variability) can yield high ICC, which might be the case in the affected (painful) arm for CDT and in both arms for PPT and PPPT.

In order to gain a more comprehensive interpretation of the measurement variability, we also calculated the SEM% (29). The SEM% varied in the affected (painful) arm from 4% to 42% and in the unaffected arm from 4% to 36% for all measurements, with lower variability for thermal detection and heat pain thresholds and higher variability for cold pain and mechanical pain thresholds. As previous studies in persons with PSSP have not reported SEM% values, our results are difficult to compare with others. However, one study has investigated the variability of thermal thresholds of the hand in healthy individuals, using the coefficient of variation in % (CV%) (33). They found CV% (log transformation of some variables) ranging from 4% to 9% in CDT, WDT and HPT, but from 85% to 86% in CPT. As CV% is comparable to SEM% (29), their results with lower variability for CDT, WDT and HPT and higher for CPT are in agreement with the results in our study.

Several factors can influence the measurement variability, for example the pain variation in itself, the QST method being used and the measurement locations that are measured. The pain variation is an inherent factor and to register identical pain thresholds on 2 occasions is unlikely (34). The variations in QST thresholds in our population might have occurred because of the somatosensory deficits or a day-to-day variation in PSSP. On the other hand, variations in QST thresholds were seen not only in the affected (painful) side, but also in the unaffected side. The magnitude of the variability was approximately the same in both arms, indicating that the QST method in itself might have influenced the result. QST is described as a psychophysical method, relying on the participant's cooperation. Two different methods, "the method of limits" and "the method of levels" can be used. The method of limits is most widely used because it is less time-consuming, which is important as the participants' concentration could be affected after stroke. The disadvantage of using "the method of limits" is that the reaction time is included in the measurements (35). The reaction time and the ability to perceive instructions might be affected after stroke, which could influence the result. However, a previous review regarding reliability of QST measurements in persons with various diagnoses reported that the 2 QST methods are equally reliable (14).

In the present study, cut-off limits of QST measures were set to 10° for cold and 50° for heat, and to 1,000 kPa for PPT and 400 g for PPPT. This is in agreement with the recommendations of QST measurements because of medical safety. A floor effect was seen in CPT; approximately half of the participants scored 10° in 1 of the 2 cold pain assessments. A ceiling effect

was also seen in the affected (painful) arm in HPT and in both arms in PPPT at the second test occasion; approximately 20% of the participants scored the maximum value. Cut-off limits of QST measures can influence the results in a test-retest situation and might mask a true QST value.

Other factors that might have influenced the measurement variability are the environment, the test procedure and the measurement locations. In our study, the environment was free from noise and distractions, a standardized test procedure was used and the examiner was unable to see the QST measurements of the first test occasion at the second test. We assessed detection and pain thresholds in predetermined measurement locations of the upper arms. However, when measuring PPT and PPPT, the perception of pain depends on exactly the point that is being measured. Wessberg et al. (36) have found that the structure and size of receptive field areas vary considerably within individuals. Thus, even if we used a standardized test protocol on both test occasions, small differences in measurement points, pressure direction or participants' sitting position could have influenced the result. Taken together, substantial variability in the QST measurements were found, but it is difficult to distinguish how much it depends on the participants or the impreciseness of the QST method.

Variations in QST thresholds have also been reported by Krassioukov et al. (37) and they recommend more than 1 QST assessment in persons with spinal cord injury when determining baseline values of pain or when evaluating effects on pain after interventions. Also, O'Neill & O'Neill (38), who investigated ways to improve the reliability of QST measurements, recommended repeated test occasions.

#### *Strengths and limitations*

A strength of the present study is that the population was adequately homogenous with regard to age and functional status and that both the affected (painful) and the unaffected upper arm were measured. An experienced examiner performed all measurements, a standardized test protocol was used, the study carefully adhered to the patient group, and the measurements were digitally recorded. The unaffected arm was tested first so the participant became familiar with the test procedure. In addition, several statistical methods were applied to evaluate the measurement variability for a group of individuals.

This study has also some limitations. The sample size was relatively small, but in agreement with other studies assessing reliability of QST measurements (15, 16). The participants were clinically assessed only on the first test occasion and a time interval of 2–3 weeks before the second test occasion was chosen for practical reasons. However, as the participants were in a stable phase post-stroke, we believe that this only has influenced the results to a small extent. Outcome measures of fatigue and cognitive impairments could also have been used. Fatigue and/or cognitive impairments might affect the QST measurements, but all participants were able to follow the test instructions and fulfilled the measurements with no problem. Participants with severe language and cognitive difficulties were not included in the present study. Therefore, the

results cannot be generalized to the entire stroke population, but, on the other hand, QST is not recommended for those individuals (39).

#### *Clinical relevance and implications for future research*

PSSP is a complex and multifactorial problem and one of the most common types of pain after stroke. To be able to evaluate changes in PSSP after interventions or over time, knowledge of the measurement variability is needed. As some of the QST variables vary widely when being repeated, several baseline assessments might be recommended, but need to be evaluated further in future studies. The study design of 1 QST measurement repeated twice revealed unacceptably low ICCs and high measurement variability. Before we can use QST routinely in clinical practice and in research, a test protocol that yields smaller measurement variability needs to be developed. In future research, larger studies are needed in order to determine how QST measurements vary for nociceptive and neuropathic PSSP and whether there is a difference in QST measurements between men and women, as suggested (13).

#### *Conclusion*

QST measurements vary in persons with PSSP, especially cold pain thresholds and mechanical pain thresholds. Before QST can be used routinely to evaluate PSSP, a test protocol with decreased variability needs to be developed.

#### ACKNOWLEDGEMENTS

The authors are grateful to the individuals who volunteered to participate. Valuable statistical advice was given by Jonas Björk, Professor of Epidemiology, Department of Occupational and Environmental Medicine, Lund University. The study was supported from the Skåne County Council Research and Development Foundation, the Swedish Stroke Association and the Promobilia Foundation.

#### REFERENCES

1. Klit H, Finnerup NB, Andersen G, Jensen TS. Central poststroke pain: a population-based study. *Pain* 2011; 152: 818–824.
2. O'Donnell MJ, Diener HC, Sacco RL, Panju AA, Vinisko R, Yusuf S, et al. Chronic pain syndromes after ischemic stroke: PROFESS trial. *Stroke* 2013; 44: 1238–1243.
3. Lindgren I, Jönsson AC, Norrving B, Lindgren A. Shoulder pain after stroke: a prospective population-based study. *Stroke* 2007; 38: 343–348.
4. Lindgren I, Lexell J, Jönsson AC, Brogårdh C. Left-sided hemiparesis, pain frequency, and decreased passive shoulder range of abduction are predictors of long-lasting poststroke shoulder pain. *PM R* 2012; 4: 561–568.
5. Gamble GE, Barberan E, Laasch HU, Bowsher D, Tyrrell PJ, Jones AK. Poststroke shoulder pain: a prospective study of the association and risk factors in 152 patients from a consecutive cohort of 205 patients presenting with stroke. *Eur J Pain* 2002; 6: 467–474.
6. Ratnasabapathy Y, Broad J, Baskett J, Pledger M, Marshall J, Bonita R. Shoulder pain in people with a stroke: a population-based study. *Clin Rehabil* 2003; 17: 304–311.
7. Isaksson M, Johansson L, Olofsson I, Eurenus E. Shoulder pain and concomitant hand oedema among stroke patients with pro-

- nounced arm paresis. *Eur J Physiother* 2013; 15: 208–214.
8. Blennerhassett JM, Gyngell K, Crean R. Reduced active control and passive range at the shoulder increase risk of shoulder pain during inpatient rehabilitation post-stroke: an observational study. *J Physiother* 2010; 56: 195–199.
  9. Roosink M, Renzenbrink GJ, Buitenweg JR, van Dongen RT, Geurts AC, Ijzerman MJ. Somatosensory symptoms and signs and conditioned pain modulation in chronic post-stroke shoulder pain. *J Pain* 2011; 12: 476–485.
  10. Soo Hoo J, Paul T, Chae J, Wilson RD. Central hypersensitivity in chronic hemiplegic shoulder pain. *Am J Phys Med Rehabil* 2013; 92: 1–13.
  11. Zeilig G, Rivel M, Weingarden H, Gaidoukov E, Defrin R. Hemiplegic shoulder pain: evidence of a neuropathic origin. *Pain* 2013; 154: 263–271.
  12. Lindgren I, Ekstrand E, Lexell J, Westergren H, Brogårdh C. Somatosensory impairments are common after stroke but have only a small impact on post-stroke shoulder pain. *J Rehabil Med* 2014; 46: 307–313.
  13. Rolke R, Magerl W, Andrews Campbell K, Schalber C, Caspari S, Birklein F, et al. Quantitative sensory testing: a comprehensive protocol for clinical trials. *Eur J Pain* 2006; 10: 77–88.
  14. Moloney NA, Hall TM, Doody CM. Reliability of thermal quantitative sensory testing: a systematic review. *J Rehabil Res Dev* 2012; 49: 191–207.
  15. Jensen-Dahm C, Werner MU, Dahl JB, Jensen TS, Ballegaard M, Hejl AM, et al. Quantitative sensory testing and pain tolerance in patients with mild to moderate Alzheimer disease compared to healthy control subjects. *Pain* 2014; 155: 1439–1445.
  16. Felix ER, Widerström-Noga EG. Reliability and validity of quantitative sensory testing in persons with spinal cord injury and neuropathic pain. *J Rehabil Res Dev* 2009; 46: 69–83.
  17. Kosek E, Ordeberg G. Lack of pressure pain modulation by heterotopic noxious conditioning stimulation in patients with painful osteoarthritis before, but not following, surgical pain relief. *Pain* 2000; 88: 69–78.
  18. Dyck PJ, Kratz KM, Lehman KA, Karnes JL, Melton LJ 3rd, O'Brien PC, et al. The Rochester Diabetic Neuropathy Study: design, criteria for types of neuropathy, selection bias, and reproducibility of neuropathic tests. *Neurology* 1991; 41: 799–807.
  19. Kemler MA, Reulen JP, van Kleef M, Barendse GA, van den Wildenberg FA, Spaans F. Thermal thresholds in complex regional pain syndrome type I: sensitivity and repeatability of the methods of limits and levels. *Clin Neurophysiol* 2000; 111: 1561–1568.
  20. Heldestad V, Linder J, Sellersjö L, Nordh E. Reproducibility and influence of test modality order on thermal perception and thermal pain thresholds in quantitative sensory testing. *Clin Neurophysiol* 2010; 121: 1878–1885.
  21. Wasner GL, Brock JA. Determinants of thermal pain thresholds in normal subjects. *Clin Neurophysiol* 2008; 119: 2389–2395.
  22. Malmström EM, Stjerna J, Högestätt ED, Westergren H. Quantitative sensory testing of temperature thresholds: Possible biomarkers for persistent pain? *J Rehabil Med* 2016; 48: 43–47.
  23. Persson AL, Brogårdh C, Sjölund BH. Tender or not tender: test-retest repeatability of pressure pain thresholds in the trapezius and deltoid muscles of healthy women. *J Rehabil Med* 2004; 36: 17–27.
  24. Pigg M, Baad-Hansen L, Svensson P, Drangsholt M, List T. Reliability of intraoral quantitative sensory testing (QST). *Pain* 2010; 148: 220–226.
  25. Huskisson EC. Measurement of pain. *Lancet* 1974; 2: 1127–1131.
  26. Carr JH, Shepherd RB, Nordholm L, Lynne D. Investigation of a new motor assessment scale for stroke patients. *Phys Ther* 1985; 65: 175–179.
  27. Barkelius K, Johansson A, Körm K, Lindmark B. [Reliability and validity testing of Modified Motor Assessment Scale according to Uppsala Akademiska hospital-95.] *Nordisk Fysioterapi* 1997; 1: 121–126 (in Swedish).
  28. Fugl-Meyer AR, Jääskö L, Leyman I, Olsson S, Steglind S. The post-stroke hemiplegic patient. I. A method for evaluation of physical performance. *Scand J Rehabil Med* 1975; 7: 13–31.
  29. Lexell JE, Downham DY. How to assess the reliability of measurements in rehabilitation. *Am J Phys Med Rehabil* 2005; 84: 719–723.
  30. Werner MU, Petersen MA, Bischoff JM. Test-retest studies in quantitative sensory testing: a critical review. *Acta Anaesthesiol Scand* 2013; 57: 957–963.
  31. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res* 2005; 19: 231–240.
  32. Bland M. *An introduction to medical statistics*. Third Edition. Oxford: Oxford University Press; 2000.
  33. Moloney NA, Hall TM, O'Sullivan TC, Doody CM. Reliability of thermal quantitative sensory testing of the hand in a cohort of young, healthy adults. *Muscle Nerve* 2011; 44: 547–552.
  34. Wylde V, Palmer S, Learmonth ID, Dieppe P. Test-retest reliability of Quantitative Sensory Testing in knee osteoarthritis and healthy participants. *Osteoarthritis Cartilage* 2011; 19: 655–658.
  35. Hansson P, Backonja M, Bouhassira D. Usefulness and limitations of quantitative sensory testing: clinical and research application in neuropathic pain states. *Pain* 2007; 129: 256–259.
  36. Wessberg J, Olausson H, Fernström KW, Vallbo AB. Receptive field properties of unmyelinated tactile afferents in the human skin. *J Neurophysiol* 2003; 89: 1567–1575.
  37. Krassioukov A, Wolfe DL, Hsieh JT, Hayes KC, Durham CE. Quantitative sensory testing in patients with incomplete spinal cord injury. *Arch Phys Med Rehabil* 1999; 80: 1258–1263.
  38. O'Neill S, O'Neill L. Improving QST reliability – more raters, tests, or occasions? A Multivariate Generalizability Study. *J Pain* 2015; 16: 454–462.
  39. Backonja MM, Attal N, Baron R, Bouhassira D, Drangsholt M, Dyck PJ, et al. Value of quantitative sensory testing in neurological and pain disorders: NeuPSIG consensus. *Pain* 2013; 154: 1807–1819.