Published in final edited form as:

J Oral Rehabil. 2015 February ; 42(2): 127–135. doi:10.1111/joor.12245.

Reliability of intraoral quantitative sensory testing (QST) in patients with atypical odontalgia and healthy controls – a multicenter study

Lene Baad-Hansen^{a,b}, Maria Pigg^{b,c,d}, Guangju Yang^{a,e}, Thomas List^{b,d,f}, Peter Svensson^{a,b}, and Mark Drangsholt^g

^aSection of Clinical Oral Physiology, Department of Dentistry, HEALTH, Aarhus University, Aarhus, Denmark

^bScandinavian Center for Orofacial Neuroscience (SCON)

^cDepartment of Endodontics, Faculty of Odontology, Malmö University, Malmö, Sweden

^dDepartment of Orofacial Pain and Jaw Function, Faculty of Odontology, Malmö University, Malmö, Sweden

^eDepartment of Prosthodontics and Center for Oral Functional Diagnosis, Treatment and Research, Peking University School of Stomatology, Beijing, China

^fDepartment of Rehabilitation Medicine, Skåne University Hospital, Lund, Sweden

^gDepartments of Oral Medicine and Oral Health Sciences, University of Washington, Seattle, WA USA

Summary

The reliability of comprehensive intraoral quantitative sensory testing (QST) protocol has not been examined systematically in patients with chronic orofacial pain. The aim of the present multicenter study was to examine test-retest and inter-examiner reliability of intraoral QST measures in terms of absolute values and z-scores as well as within-session coefficients of variation (CV) values in patients with atypical odontalgia (AO) and healthy pain-free controls. Forty-five AO patients and 68 healthy controls were subjected to bilateral intraoral gingival QST and unilateral extratrigeminal QST (thenar) on three occasions (twice on one day by two different examiners and once approximately one week later by one of the examiners). Intraclass correlation coefficients and kappa values for inter-examiner and test-retest reliability were computed. Most of the standardized intraoral QST measures showed fair to excellent inter-examiner (9–12 of 13 measures) and test-retest (7–11 of 13 measures) reliability. Furthermore, no robust differences in reliability measures or within-session variability (CV) were detected between AO patients and the healthy reference group. These reliability results in chronic orofacial pain patients support earlier suggestions based on data from healthy subjects that intraoral QST is sufficiently reliable for use

Correspondence: Dr. Lene Baad-Hansen, Section of Clinical Oral Physiology, Department of Dentistry, HEALTH, Aarhus University, Vennelyst Boulevard 9, DK-8000 Aarhus C, Denmark, lene.hansen@odontologi.au.dk.

All authors declare no conflicts of interest.

as a part of a comprehensive evaluation of patients with somatosensory disturbances or neuropathic pain in the trigeminal region.

Keywords

Quantitative sensory testing; Somatosensory testing; Reliability; Orofacial pain; Chronic intraoral pain; Neuropathic pain; Trigeminal nerve

Background

Quantitative sensory testing (QST) is used for assessment of somatosensory function (1,2). A comprehensive standardized QST protocol was presented in 2006 and has been adapted for intraoral use (1,3–5) and is considered a valuable tool in neuropathic pain diagnosis (6). Utilizing the standardized QST protocol, somatosensory abnormalities were found in 92% of neuropathic pain patients (3), 87% of patients with atypical odontalgia (AO) and 96 % of patients with peripheral painful traumatic trigeminal neuropathy (5,7,8). The individual QST scores are summarized as somatosensory z-score profiles to give a quick overview of any somatosensory loss or gain (3–5). Hence, knowledge about possible pain mechanisms in each patient can be obtained and individualized therapy may potentially be instituted (4).

AO is a chronic pain condition located in teeth and jaws and has been proposed to involve neuropathic pain mechanisms possibly caused by deafferentation of small-order trigeminal neurons (9,10–14). However, some AO cases may also be suffering from “functional” pain due to abnormally increased responsiveness of the sensory apparatus (15). Differentiation of AO pain from inflammatory tooth pain conditions is often difficult because of similar clinical presentations but the appropriate management differs radically between conditions (1). The term AO has been widely discussed and it has been suggested to rename the condition painful traumatic trigeminal neuropathy (8) or persistent dentoalveolar pain (PDAP)(16). Despite short-comings in the definition of AO, this was indeed the term used in the present study because the patients were included and tested before updated terms were published.

Reliability of QST is affected by many factors and relies on standardized test protocols, patient instructions and trained examiners (3,4). A multicenter study of neuropathy cases and matched controls reported good reliability on cutaneous sites (1). Intraorally, the standardized QST protocol is applicable with minor adjustments with acceptable reliability in pain-free subjects (17). However, the reliability of intraoral QST measures in chronic orofacial pain patients and the reliability of the somatosensory z-scores has not yet been reported.

The within-session coefficient of variation (CV) of QST measures is higher for some parameters in patients with painful temporomandibular disorders (TMD) than in healthy participants (18). Estimation of such within-session variability between three or five repeated threshold measurements within a few minutes has been suggested as a possible additional valuable piece of information (18).

The aims of the present multi-center study were to examine test-retest and inter-examiner reliability of intraoral QST measures (absolute and z-scores) and to compare within-session CV values between AO patients and healthy controls.

Materials and methods

Participants

The present part of a multi-center study involved three centers; Malmö University, Sweden (MU); University of Washington, USA (UW) and Aarhus University, Denmark (AU) (5).

Forty-five AO patients (mean age 56 ± 13 years, 38 women, 7 men) and 68 pain-free controls (mean age 42 ± 15 years, 42 women, 26 men) completed the reliability part of the present study. The AO cases were recruited among patients in the orofacial pain clinics of each center after a thorough clinical and radiographic examination (5). The controls were responders to advertisements without prior knowledge of QST. The inclusion criteria for the AO group was persistent pain located in a tooth or a region where a tooth was extracted, with no signs of pathology (5,7,10,13,19). The diagnostic assessment has been described in detail elsewhere (5). Exclusion criteria for AO patients were presence of other known orofacial pain conditions (odontogenic pain; primary headaches; medical, psychiatric or personality disorders). AO patients with comorbid TMD were not excluded as long as they could clearly distinguish between their two pain conditions (5,20).

Inclusion criteria for the control group was age >18 years, good health with no orofacial pain complaints, and exclusion criteria were serious dental, medical, psychiatric or personality disorders (5).

In agreement with the Declaration of Helsinki, local ethics committees approved the study. All participants were asked to sign an informed consent form designed according to local ethical requirements.

Test procedure

The buccal attached gingiva of the painful tooth was chosen as test area. In controls the gingiva adjacent to tooth 23–24 was chosen as the test area. The control area was the contralateral “mirror-image” site. The extra-trigeminal control area was the skin overlying the thenar eminence of the patient’s right hand. Testing of one area took ~30 min and a complete test at all three areas took ~2 h to perform. Testing order was extra-trigeminal control area first, followed by control area and test area for all trials. All examinations were performed with the subject supine or semi-supine in a quiet room with even temperature. All reusable QST devices were cleaned with disinfection wipes, immersed in 1% Chlorhexidine solution, or sterilized in chemical autoclaves between participants and all single-use items were discarded.

A total of 12 trained examiners executed the tests (AU:2; MU:2; UW:8). All examiners were either previously experienced in intraoral QST or had completed pre-study training sessions. Subjects received the same standardized instructions at all three examinations.

To calculate the inter-examiner reliability, each participant was tested twice on the same day by two different examiners. To calculate intra-examiner (test-retest) reliability, the testing was repeated after approximately one week by one of the examiners. On each occasion, the examiner was blinded to the results of any previous examination.

Test stimuli

A full QST according to the standardized protocol was performed: cold and warmth detection thresholds (CDT, WDT), thermal sensory limen (TSL), paradoxical heat sensations (PHS), cold and heat pain thresholds (CPT, HPT), mechanical detection threshold (MDT), mechanical pain threshold (MPT), mechanical pain sensitivity (MPS), dynamic mechanical allodynia (DMA), vibration detection threshold (VDT) and pressure pain threshold (PPT) (4,5,17). For a detailed description, see supplementary material S1 and (5).

Statistical analyses

Absolute values have been reported elsewhere (5). PHS and DMA (discrete variables) were analyzed separately. CPT, HPT and VDT were normally distributed. All other parameters were log-transformed before analysis (3,4). A z-transformation of each variable was performed based on the control group data with adjustment of sign in such a way that positive z-scores indicated gain of somatosensory function and negative z-scores indicated loss of function (3–5). The individual test area z-scores were calculated as $(\text{mean}_{\text{reference}} - \text{individual value})/\text{SD}_{\text{reference}}$. The z-scores from the present patients have been reported earlier (5).

Differences in age and gender distribution between groups were analyzed using unpaired t-test and χ^2 -test. Since age- and gender-matching was not obtained using all the included healthy participants, two separate analyses of inter- and intra-examiner reliability were performed; one including all 68 control participants and another analysis, where the 25 youngest control participants (9F and 16M) were excluded from analysis. After such exclusion, there were no statistically significant age- and gender-differences between groups ($P>0.097$).

Inter- and intra-examiner reliability was analyzed using the intraclass correlation coefficient (ICC) for quantitative variables (absolute values and z-scores) and Cohen's kappa for PHS. ICC<0.4 was considered poor agreement; 0.4–0.59 fair; 0.6–0.75 good; and >0.75 excellent (21). For kappa, values 0.2 were considered poor agreement; 0.21–0.40 fair; 0.41–0.60 moderate; 0.61–0.80 good; and 0.81–1.00 excellent (22). To examine possible differences in reliability between centers, ICC values of z-scores in healthy participants were computed for each center. As an overall measure of inter-examiner and intra-examiner measurement agreement, the standard deviations (SD) of the differences between measured values at the test area of each QST parameter in each session were computed: Inter-examiner measurement agreement was computed as the difference between the values obtained by examiner 2 and examiner 1; Intra-examiner measurement agreement was computed as the difference between the two values obtained by examiner 1 on separate days. These SD values will be presented in a supplementary table.

Within-session CV values (the ratio of the standard deviation to the mean of the triplicate measure) of each individual QST parameter (18). Since the TSL was the difference limen for alternating cold and warm stimuli, three limens were calculated based on three alternating cold and warmth detections. MPS or DMA in the individual test was calculated as the geometric mean of numerical ratings for pinprick or tactile stimuli. CV values were calculated for test area only. CV values of the five “Yes” responses and five “No” responses of MDT and MPT were calculated separately as “MDT-Y”, “MDT-N”, “MPT-Y”, and “MPT-N”. CV values were compared between AO patients and healthy controls with unpaired t-tests. Two comparisons were made between groups, one with the total number of healthy controls (n=68) and one with the matched sub-group of healthy controls (n=43).

Statistical tests were performed two-tailed and at the 5% significance level.

Results

For information about absolute values of QST parameters, please refer to (5).

ICC and kappa values for inter-examiner and test-retest reliability of QST measures from all test sites are given in Table 1. All reliability measures are given for AO patients (n=45) as well as all included healthy controls in the present part of the study (n=68) and after obtaining age- and gender-matching by removing the 25 youngest controls (n=43: in *italics* in table 1). The adjustment of the control group to obtain age- and gender-matching with the patient group did not have a systematic influence the reliability measures (Table 1). The majority of inter-examiner ICC values were in the good to excellent range (Table 1). All reliability measures were similar between groups.

An example of repeated QST z-score profiles is shown in Fig. 1. ICCs for inter-examiner and test-retest reliability of QST z-scores for the test area and extra-trigeminal control area are shown in Table 2 for both groups. The inter- and intra-examiner reliability of the z-scores (Table 2) were in the same range as for the absolute values of the QST measures (Table 1).

ICCs for z-scores of continuous intraoral QST variables for each center are shown in Table 3.

Inter-examiner and intra-examiner measurement agreement values (SD of differences between examinations) are shown in a supplementary table. All QST variables show similar measurement agreement between AO patients and healthy participants, except for MDT, which displays a considerably larger SD (poorer agreement) in the AO group (inter-examiner: 92.6 mN; intra-examiner: 100.3 mN) compared with the healthy group (inter-examiner: 28.2 mN; intra-examiner: 24.3 mN).

The CV values for AO patients and the total number of healthy controls are shown in Table 4. Only the tactile cotton wisp stimulation (part of the allodynia evaluation) showed higher CV values in AO patients than in controls (P=0.037). Otherwise, no statistically significant differences in CV values were detected between patients and controls (n=68 or n=43).

Discussion

Reliability of intraoral QST

This is the first study on reliability of a standardized battery of intraoral QST measures obtained in accordance with the latest guidelines (2) in chronic orofacial pain patients. The main finding of this study was that reliability of intraoral QST absolute values and z-scores was similar between patients with AO and healthy controls, regardless of whether the healthy group was age- and gender-matched or not. For most parameters, inter- and intra-examiner reliability was at least fair in both groups. This is in accordance with our earlier reliability study performed in 21 healthy participants (17). The present study included a considerably larger group of healthy participants (n=68) as well as 45 AO patients. Since the 68 healthy participants originally included in the reliability part of this multi-centre study was not properly age- and gender-matched to the AO patient population, a separate analysis was performed after exclusion of the 25 youngest healthy participants to control for age- or gender-effects on the reliability. QST measures are to some extent dependent on vigilance and reaction time and age may be an important confounding factor in such reliability evaluations (2). However, similar ICC and kappa values were found between the total group of healthy participants and the matched group in the present study. Moreover, also measurement agreement was similar between groups, except for mechanical detection thresholds (MDT), where poorer measurement agreements were found in AO patients compared with healthy controls.

No other studies have examined ICC values of standardized QST in both patient populations and healthy matched controls. However, the reliability of QST at the affected skin site has been compared with the non-affected side in neuropathic pain patients and correlations were stronger between examinations at the affected site than at the non-affected site (1). This may be due to a larger range of values obtained in patients in the affected region compared with the non-affected region. If the present ICC values are compared with the reliability study using QST on skin (1), they may seem comparably low. However, in the study by Geber et al., reliability was evaluated using Pearson's and Spearman correlation analyses, which measures the strength of the association between measures but does not take into account the agreement between values (1,23). A risk of a slight overestimation of reliability should therefore be considered, when comparisons across studies are attempted (23). Nevertheless, due to the special anatomical challenges for intraoral QST performance, it is indeed believed to be more variable than QST on skin (2,17). On the other hand, many parts of the trigeminal region are densely innervated, which could be hypothesized to decrease the variability. However, from the present data it can be seen that the reliability was generally higher at the extra-trigeminal site than for the intraoral sites in both groups.

Reliability of specific intraoral QST measures and z-scores

Across groups and test sites, PHS had the poorest reliability, which is in accordance with the Geber et al. study on skin (1). In the patient population, the reliability of the DMA measure shows higher reliability scores than in the healthy group. However, in healthy controls, DMA scores are very often '0', which results in ICC values that often cannot be calculated or are not meaningful (17). From the present reliability study in a fairly large study sample,

no single QST measure stands out as the most reliable. However, across groups, sites and examinations, TSL, CPT, MPT, VDT and PPT shows the highest reliability. Therefore, from a reliability point of view these measures may prove the most useful in studies including follow-up on somatosensory sensitivity in a population.

Another important finding was that the inter- and intra-examiner reliability of the z-scores was in the same range as for the absolute QST values. This is the first study ever to report the reliability of QST z-scores. Z-scores are calculated as the difference between the individual QST parameter and the mean value of the healthy reference group divided by the standard deviation of the healthy reference group (3–5). Mathematically this means that the reliability of the z-scores is very closely linked to the reliability of the absolute QST scores, since the means and SDs of the reference group scores between sessions are less variable.

Within-session variability

The final finding of the present study was that CV values of within-session repeated measurements were also similar between patients and controls. Only the pain score to cotton wisp stimulation showed significantly higher CV values in AO patients than in healthy controls. In healthy controls, such stimulation was of course almost never reported as painful and, therefore, the CV values were very low. The overall similarity in CV values between groups in the present study is in contrast to recent findings of CV values of facial QST between patients with painful TMD and healthy controls, where CV values for several measures were larger in TMD patients than in controls (18). Based on the present results, only cotton wisp stimulation CV scores seem to have the possibility to distinguish between AO patients and controls. Regarding within-session variability, WDT, HPT and VDT measures show the lowest CVs, i.e. they are very stable measures when repeated in triplicates.

Implications for diagnosis

It is clear that QST examinations cannot stand alone and should be combined with a thorough patient history and preferably also with relevant neurophysiological tests or imaging techniques, whenever available, to increase diagnostic sensitivity and specificity (2,6,24). Based on the QST findings from our previously published part of the study (5), a large proportion (87%) of the AO patients included in the study could eventually have been labelled “probable neuropathic pain” patients. However, a diagnosis of “definite neuropathic pain” would demand further confirmatory tests, such as neurophysiological tests (6). The lack of QST availability in primary care is a great challenge due to a risk of misdiagnosis and, consequently, improper treatment. Nevertheless, preliminary qualitative sensory testing can easily be performed everywhere with very simple instruments before referral to specialized clinics (20).

Conclusions

Most standardized intraoral QST measures showed fair to excellent inter-examiner and test-retest reliability. Furthermore, no robust differences in reliability or within-session variability were detected between AO patients and healthy controls.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The present study was financially supported by the Faculty of Odontology, Malmö University and National Institutes of Health grants K12 DE14069 and R21-DE018768.

References

1. Geber C, Klein T, Azad S, Birklein F, Gierthmühlen J, Hüge V, Lauchart M, Nitzsche D, Stengel M, Valet M, Baron R, Maier C, Tolle T, Treede RD. Test-retest and interobserver reliability of quantitative sensory testing according to the protocol of the German Research Network on Neuropathic Pain (DFNS): a multi-centre study. *Pain*. 2011; 152:548–556. [PubMed: 21237569]
2. Svensson P, Baad-Hansen L, Pigg M, List T, Eliav E, Ettlin D, Michelotti A, Tsukiyama Y, Matsuka Y, Jääskeläinen SK, Essick G, Greenspan JD, Drangsholt M. Guidelines and recommendations for assessment of somatosensory function in oro-facial pain conditions - a taskforce report. *J Oral Rehabil*. 2011; 38:366–394. [PubMed: 21241350]
3. Maier C, Baron R, Tolle TR, Binder A, Birbaumer N, Birklein F, Gierthmühlen J, Flor H, Geber C, Hüge V, Krumova EK, Landwehrmeyer GB, Magerl W, Maihofner C, Richter H, Rolke R, Scherens A, Schwarz A, Sommer C, Tronnier V, Uceyler N, Valet M, Wasner G, Treede RD. Quantitative sensory testing in the German Research Network on Neuropathic Pain (DFNS): somatosensory abnormalities in 1236 patients with different neuropathic pain syndromes. *Pain*. 2010; 150:439–450. [PubMed: 20627413]
4. Rolke R, Magerl W, Campbell KA, Schalber C, Caspari S, Birklein F, Treede RD. Quantitative sensory testing: a comprehensive protocol for clinical trials. *Eur J Pain*. 2006; 10:77–88. [PubMed: 16291301]
5. Baad-Hansen L, Pigg M, Ivanovic SE, Faris H, List T, Drangsholt M, Svensson P. Intraoral somatosensory abnormalities in patients with atypical odontalgia—a controlled multicenter quantitative sensory testing study. *Pain*. 2013; 154:1287–1294. [PubMed: 23725780]
6. Treede RD, Jensen TS, Campbell JN, Cruccu G, Dostrovsky JO, Griffin JW, Hansson P, Hughes R, Nurmikko T, Serra J. Neuropathic pain. Redefinition and a grading system for clinical and research purposes. *Neurology*. 2008; 70:1630–1635. [PubMed: 18003941]
7. List T, Leijon G, Svensson P. Somatosensory abnormalities in atypical odontalgia: A case-control study. *Pain*. 2008; 139:333–341. [PubMed: 18571324]
8. Benoliel R, Zadik Y, Eliav E, Sharav Y. Peripheral painful traumatic trigeminal neuropathy: clinical features in 91 cases and proposal of novel diagnostic criteria. *J Orofac Pain*. 2012; 26:49–58. [PubMed: 22292140]
9. Nixdorf DR, Moana-Filho EJ, Law AS, McGuire LA, Hodges JS, John MT. Frequency of nonodontogenic pain after endodontic therapy: a systematic review and meta-analysis. *J Endod*. 2010; 36:1494–1498. [PubMed: 20728716]
10. Baad-Hansen L. Atypical odontalgia - pathophysiology and clinical management. *J Oral Rehabil*. 2008; 35:1–11. [PubMed: 18190356]
11. Baad-Hansen L, Juhl GI, Jensen TS, Brandsborg B, Svensson P. Differential effect of intravenous S-ketamine and fentanyl on atypical odontalgia and capsaicin-evoked pain. *Pain*. 2007; 129:46–54. [PubMed: 17088020]
12. Baad-Hansen L, List T, Kaube H, Jensen TS, Svensson P. Blink reflexes in patients with atypical odontalgia and matched healthy controls. *Exp Brain Res*. 2006; 172:498–506. [PubMed: 16489436]
13. Melis M, Lobo SL, Ceneviz C, Zawawi K, Al Badawi E, Maloney G, Mehta N. Atypical odontalgia: a review of the literature. *Headache*. 2003; 43:1060–1074. [PubMed: 14629241]
14. Woda A, Pionchon P. A unified concept of idiopathic orofacial pain: pathophysiologic features. *J Orofac Pain*. 2000; 14:196–212. [PubMed: 11203755]

15. Woolf CJ. Pain: moving from symptom control toward mechanism-specific pharmacologic management. *Ann Intern Med.* 2004; 140:441–451. [PubMed: 15023710]
16. Nixdorf DR, Drangsholt MT, Ettlin DA, Gaul C, De Leeuw R, Svensson P, Zakrzewska JM, DE Laat A, Ceusters W. Classifying orofacial pains: a new proposal of taxonomy based on ontology. *J Oral Rehabil.* 2012; 39:161–169. [PubMed: 21848527]
17. Pigg M, Baad-Hansen L, Svensson P, Drangsholt M, List T. Reliability of intraoral quantitative sensory testing (QST). *Pain.* 2010; 148:220–226. [PubMed: 20022428]
18. Yang G, Baad-Hansen L, Wang K, Xie QF, Svensson P. A study on variability of quantitative sensory testing in healthy participants and painful temporomandibular disorder patients. *Somatosens Mot Res.* 2014 In Press.
19. List T, Leijon G, Helkimo M, Öster A, Svensson P. Effect of local anesthesia on atypical odontalgia—a randomized controlled trial. *Pain.* 2006; 122:306–314. [PubMed: 16564621]
20. Dworkin SF, LeResche L. Research diagnostic criteria for temporomandibular disorders: review, criteria, examinations and specifications, critique. *J Craniomandib Disord.* 1992; 6:301–355. [PubMed: 1298767]
21. Baad-Hansen L, Pigg M, Ivanovic SE, Faris H, List T, Drangsholt M, Svensson P. Chairside intraoral qualitative somatosensory testing: reliability and comparison between patients with atypical odontalgia and healthy controls. *J Orofac Pain.* 2013; 27:165–170. [PubMed: 23630688]
22. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull.* 1979; 86:420–428. [PubMed: 18839484]
23. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977; 33:159–174. [PubMed: 843571]
24. Muller R, Buttner P. A critical discussion of intraclass correlation coefficients. *Stat Med.* 1994; 13:2465–2476. [PubMed: 7701147]
25. Jääskeläinen SK. Clinical neurophysiology and quantitative sensory testing in the investigation of orofacial pain and sensory function. *J Orofac Pain.* 2004; 18:85–107. [PubMed: 15250429]

Fig. 1.

An example of three z-score profiles obtained in a patient with atypical odontalgia (AO) from each of the three QST examinations during the study. Similar z-score profile patterns are seen across the three examinations (1a, 1b and 2a). Examination 1a and 1b were performed by different examiners on the same day. Examination 2a was performed on a separate day. For further information about absolute values and z-scores of quantitative sensory testing (QST) parameters please refer to (5).

Table 1

Inter-examiner (**1a**) and test-retest (**1b**) reliability (interclass correlations (ICC) for continuous variables and kappa values for categorical variables) for absolute quantitative sensory testing (QST) variables in patients with atypical odontalgia (AO) (AO, n=45), healthy control subjects (Controls, n=68) and both groups together (Total, n=113) for intraoral pain site (test area: TA), intraoral control area (CA), and extra-trigeminal control site on hand (ECA). (For CS group without pain, TA corresponds to left side and CA to right side).

1a. Inter-examiner reliability									
AO		Controls				Total			
TA	CA	ECA	TA	CA	ECA	TA	CA	ECA	
CDT	0.53	0.50	0.42	0.50 (0.49)	0.53 (0.61)	0.82 (0.72)	0.51 (0.53)	0.52 (0.59)	0.55 (0.48)
WDT	0.34	0.62	0.72	0.64 (0.79)	0.56 (0.57)	0.76 (0.86)	0.46 (0.45)	0.58 (0.61)	0.74 (0.78)
TSL	0.68	0.29	0.74	0.66 (0.79)	0.65 (0.71)	0.83 (0.87)	0.67 (0.74)	0.50 (0.51)	0.78 (0.78)
PHS	0.04	0.30	0.08	0.14 (0.27)	0.27 (0.52)	-0.06 (0.00)	0.11 (0.25)	0.30 (0.50)	0.00 (0.15)
CPT	0.78	0.76	0.67	0.59 (0.63)	0.57 (0.64)	0.61 (0.65)	0.69 (0.72)	0.64 (0.68)	0.64 (0.67)
HPT	0.50	0.32	0.65	0.44 (0.57)	0.44 (0.50)	0.77 (0.76)	0.48 (0.51)	0.38 (0.37)	0.71 (0.70)
MDT	0.44	0.44	0.82	0.41 (0.42)	0.55 (0.51)	0.52 (0.52)	0.45 (0.44)	0.48 (0.46)	0.81 (0.81)
MPT	0.74	0.60	0.71	0.58 (0.50)	0.63 (0.55)	0.60 (0.62)	0.65 (0.63)	0.62 (0.59)	0.66 (0.67)
MPS	0.62	0.76	0.40	0.56 (0.39)	0.68 (0.94)	0.71 (0.68)	0.60 (0.60)	0.70 (0.83)	0.58 (0.52)
DMA	0.79	0.83	0.71	0.64 (0.99)	0.47 (0.84)	0.93 (0.78)	0.78 (0.80)	0.72 (0.83)	0.73 (0.71)
WUR	0.66	0.13	0.57	0.44 (0.50)	0.52 (0.52)	0.42 (0.44)	0.57 (0.63)	0.29 (0.26)	0.49 (0.49)
VDT	0.60	0.46	0.63	0.49 (0.35)	0.60 (0.50)	0.67 (0.63)	0.54 (0.49)	0.56 (0.54)	0.64 (0.63)
PPT	0.66	0.45	0.60	0.50 (0.51)	0.43 (0.39)	0.60 (0.60)	0.60 (0.62)	0.44 (0.43)	0.59 (0.58)

1b. Test-retest reliability									
AO		Controls				Total			
TA	CA	ECA	TA	CA	ECA	TA	CA	ECA	
CDT	0.50	0.48	0.38	0.60 (0.66)	0.46 (0.59)	0.59 (0.59)	0.55 (0.56)	0.47 (0.52)	0.46 (0.43)
WDT	0.52	0.45	0.70	0.67 (0.71)	0.53 (0.60)	0.83 (0.94)	0.59 (0.56)	0.51 (0.53)	0.77 (0.80)
TSL	0.72	0.47	0.75	0.44 (0.58)	0.66 (0.83)	0.79 (0.80)	0.57 (0.65)	0.59 (0.63)	0.77 (0.76)
PHS	0.27	0.09	0.36	0.02 (0.18)	0.24 (0.38)	0.20 (-0.05)	0.18 (0.27)	0.19 (0.33)	0.27 (0.41)
CPT	0.51	0.41	0.66	0.55 (0.52)	0.60 (0.62)	0.36 (0.46)	0.53 (0.49)	0.52 (0.48)	0.51 (0.56)
HPT	0.32	0.31	0.66	0.49 (0.66)	0.42 (0.55)	0.49 (0.49)	0.42 (0.41)	0.37 (0.41)	0.55 (0.55)

1b. Test-retest reliability

AO	Controls						Total		
	TA	CA	ECA	TA	CA	ECA	TA	CA	ECA
MDT	0.13	0.25	0.62	0.63 (0.60)	0.43 (0.35)	0.47 (0.58)	0.25 (0.22)	0.32 (0.29)	0.63 (0.63)
MPT	0.62	0.51	0.68	0.62 (0.58)	0.63 (0.64)	0.54 (0.53)	0.63 (0.61)	0.59 (0.58)	0.61 (0.63)
MPS	0.17	0.41	0.58	0.71 (0.97)	0.28 (0.40)	0.44 (0.58)	0.48 (0.47)	0.30 (0.40)	0.48 (0.57)
DMA	0.13	0.36	0.52	– (–)	–0.03 (0.01)	–0.04 (–)	0.14 (0.15)	0.29 (0.37)	0.41 (0.52)
WUR	0.04	0.63	0.31	0.23 (0.17)	0.24 (0.02)	0.18 (0.22)	0.07 (0.05)	0.53 (0.57)	0.24 (0.27)
VDT	0.62	0.55	0.67	0.64 (0.55)	0.64 (0.60)	0.58 (0.53)	0.64 (0.58)	0.62 (0.57)	0.62 (0.58)
PPT	0.67	0.51	0.64	0.64 (0.75)	0.65 (0.78)	0.32 (0.35)	0.66 (0.71)	0.59 (0.63)	0.50 (0.51)

Numbers in *italics* represent ICC and kappa values obtained using only 43 age- and gender-matched controls subjects. Numbers in **bold** represent fair to excellent reliability (ICCs 0.4). CDT: cold detection threshold, WDT: warmth detection threshold, TSL: thermal sensory limen, PHS: paradoxical heat sensation, CPT: cold pain threshold, HPT: heat pain threshold, MDT: mechanical detection threshold, MPT: mechanical pain threshold, MPS: mechanical pain sensitivity, DMA: dynamic mechanical allodynia, WUR: wind-up ratio, VDT: vibration detection threshold, PPT: pressure pain threshold.

Table 2

Intraclass correlation coefficients (ICC) of individual z-scores of quantitative sensory testing (QST) parameters (log-transformed if not normally distributed without transformation) for the intraoral test area in patients with atypical odontalgia (AO) (n=45) and healthy control participants (Controls) (n=68).

	AO		Controls	
	Interexaminer	Test-retest	Interexaminer	Test-retest
LogCDT	0.58	0.48	0.46	0.60
LogWDT	0.45	0.52	0.62	0.62
LogTSL	0.72	0.60	0.58	0.48
CPT	0.76	0.54	0.59	0.59
HPT	0.40	0.44	0.44	0.53
logMDT	0.59	0.51	0.62	0.72
LogMPT	0.74	0.61	0.58	0.57
LogMPS	0.39	0.39	0.81	0.83
LogDMA	0.62	0.76	0.83	0.83
LogWUR	0.59	0.25	0.54	0.60
VDT	0.61	0.45	0.52	0.64
LogPPT	0.78	0.73	0.52	0.74

Numbers in **bold** represent fair to excellent reliability (ICCs ≥ 0.4). CDT: cold detection threshold, WDT: warmth detection threshold, TSL: thermal sensory limen, PHS: paradoxical heat sensation, CPT: cold pain threshold, HPT: heat pain threshold, MDT: mechanical detection threshold, MPT: mechanical pain threshold, MPS: mechanical pain sensitivity, DMA: dynamic mechanical allodynia, WUR: wind-up ratio, VDT: vibration detection threshold, PPT: pressure pain threshold.

Table 3

Intraclass correlation coefficients (ICC) of individual z-scores of quantitative sensory testing (QST) parameters (log-transformed if not normally distributed without transformation) for the intraoral test area in healthy participants between the three study centers: University of Washington (UW: n = 24), Malmö University (MU: n = 33) and Aarhus University (AU: n = 11).

	UW	UW	MU	MU	AU	AU
	Interexaminer	Test-retest	Interexaminer	Test-retest	Interexaminer	Test-retest
LogCDT	0.25	0.34	0.28	0.61	0.53	0.65
LogWDT	0.41	0.43	0.45	0.42	0.56	0.50
LogTSL	0.21	0.05	0.58	0.35	0.79	0.74
CPT	0.37	0.44	0.60	0.39	0.35	0.73
HPT	0.34	0.52	0.48	0.48	0.88	0.85
logMDT	0.56	0.89	0.68	0.74	0.69	0.26
LogMPT	0.47	0.65	0.71	0.61	0.52	0.33
LogMPS	0.79	0.86	0.67	0.65	0.90	0.84
LogDMA	0.59	0.56	-	-	0.96	0.94
LogWUR	0.54	0.49	0.29	0.61	0.65	0.79
VDT	0.52	0.56	0.49	0.75	0.56	0.38
LogPPT	0.31	0.74	0.73	0.82	0.46	0.54

Numbers in **bold** represent fair to excellent reliability (ICCs ≥ 0.4). CDT: cold detection threshold, WDT: warmth detection threshold, TSL: thermal sensory linen, PHS: paradoxical heat sensation, CPT: cold pain threshold, HPT: heat pain threshold, MDT: mechanical detection threshold, MPT: mechanical pain threshold, MPS: mechanical pain sensitivity, DMA: dynamic mechanical allodynia, WUR: wind-up ratio, VDT: vibration detection threshold, PPT: pressure pain threshold.

Table 4

The coefficients of variance (CV) of within-session repeated measurements of intraoral quantitative sensory testing (QST) measures in patients with atypical odontalgia (AO) and healthy control participants.

CV of parameter	Controls		AO		P
	Mean (%)	SD	Mean (%)	SD	
CDT	22.2	40.8	17.6	22.8	0.815
WDT	2.7	3.0	2.9	3.5	0.575
TSL	20.4	17.8	20.1	19.2	0.323
CPT	65.3	198.0	40.8	48.5	0.371
HPT	2.2	2.8	2.5	3.3	0.795
MDT-Y	43.4	25.0	45.6	18.4	0.084
MDT-N	41.4	24.5	43.9	26.6	0.960
MPT-Y	41.0	28.4	43.5	27.8	0.102
MPT-N	41.2	26.5	42.5	25.0	0.981
MPS	45.0	24.0	49.6	26.2	0.251
CW	2.8	22.9	11.3	40.1	0.037*
QT	6.4	30.8	12.4	37.9	0.210
BR	21.6	55.7	12.3	38.8	0.686
WUR	36.6	30.6	38.3	24.4	0.104
VDT	4.0	3.5	3.5	3.3	0.783
PPT	12.1	7.0	14.2	10.4	0.720

CDT: cold detection threshold, WDT: warmth detection threshold, TSL: thermal sensory limen, PHS: paradoxical heat sensation, CPT: cold pain threshold, HPT: heat pain threshold, MDT-Y: mechanical detection threshold (yes), MDT-N: mechanical detection threshold (no), MPT-Y: mechanical pain threshold (yes), MPT-N: mechanical pain threshold (no), MPS: mechanical pain sensitivity, CW: cotton wisp, QT: cotton wool Q-tip, BR: brush, WUR: wind-up ratio, VDT: vibration detection threshold, PPT: pressure pain threshold.

* indicates statistically significant difference between AO patients and healthy participants (CS)