Eur J Appl Physiol (2011) 111:83–92
DOI 10.1007/s00421-010-1634-0

ORIGINAL ARTICLE

Test–retest reliability of the nociceptive withdrawal reflex and electrical pain thresholds after single and repeated stimulation in patients with chronic low back pain

José A. Biurrun Manresa · Alban Y. Neziri · Michele Curatolo · Lars Arendt-Nielsen · Ole K. Andersen

Accepted: 23 August 2010 / Published online: 3 September 2010

Abstract Recent studies have shown that the nociceptive withdrawal reflex threshold (NWR-T) and the electrical pain threshold (EP-T) are reliable measures in pain-free populations. However, it is necessary to investigate the reliability of these measures in patients with chronic pain in order to translate these techniques from laboratory to clinic. The aims of this study were to determine the test–retest reliability of the NWR-T and EP-T after single and repeated (temporal summation) electrical stimulation in a group of patients with chronic low back pain, and to investigate the association between the NWR-T and the EP-T. To this end, 25 patients with chronic pain participated in three identical sessions, separated by 1 week in average, in which the NWR-T and the EP-T to single and repeated stimulation were measured. Test–retest reliability was assessed using intra-class correlation coefficient (ICC), coefficient of variation (CV), and Bland–Altman analysis. The association between the thresholds was assessed using the coefficient of determination (r^2). The results showed good-to-excellent reliability for both NWR-T and EP-T in all cases, with average ICC values ranging 0.76–0.90 and average CV values ranging 12.0–17.7%. The association between thresholds was better after repeated stimulation than after single stimulation, with average r^2 values of 0.83 and 0.56, respectively. In conclusion, the NWR-T and the EP-T are reliable assessment tools for assessing the sensitivity of spinal nociceptive pathways in patients with chronic pain.

Keywords Nociceptive withdrawal reflex threshold · Electrical pain threshold · Reliability · Chronic low back pain · Experimental pain · Temporal summation

Communicated by Fausto Baldissera.

J. A. Biurrun Manresa (✉) · L. Arendt-Nielsen · O. K. Andersen
Integrative Neuroscience Group,
Center for Sensory-Motor Interaction,
Department of Health Science and Technology,
Aalborg University, Fredrik Bajers Vej 7,
9220 Aalborg Øst, Denmark
e-mail: jbiurrun@hst.aau.dk

A. Y. Neziri · M. Curatolo
University Department of Anesthesiology and Pain Therapy,
University Hospital of Bern, Inselspital,
Freiburgstrasse, 3010 Bern, Switzerland

Introduction

The nociceptive withdrawal reflex (NWR) is a typical reaction observed in almost all living species, with the purpose of withdrawing the body from potential tissue-damaging agents (Clarke and Harris 2004). This reflex was originally named 'flexion reflex' (Sherrington 1910), although later research showed that an extension reflex could also be elicited (Hagbarth 1960) when an extremity was stimulated, thus expanding the concept to the more general term 'withdrawal reflex'.

A NWR can be elicited by natural and artificial stimuli. Examples of natural stimuli are heat and punctuate pinprick, which activate specific nociceptors in the skin (Schouenborg and Kalliomäki 1990). Electrical stimulation is the most widely used artificial method for eliciting the NWR in humans (Tørring et al. 1981). This stimulus bypasses the skin receptor and generates synchronous action potentials in multiple sensory nerves, consequently evoking a synchronized reflex. Moreover, it has been shown that repeated electrical stimulation can induce temporal summation of the NWR (Arendt-Nielsen et al. 1994), thus extending the potential of the NWR to the assessment of central pain processes (Andersen et al. 2005; Serrao et al. 2004).

The NWR has also been proven useful in the assessment of physiological, chemical, and pharmacological modulation of nociceptive transmission/processing and for investigations in patients with chronic pain (for a review see Sandrini et al. 2005). Patients with chronic pain syndromes such as whiplash (Curatolo et al. 2001), fibromyalgia (Graven-Nielsen et al. 1999), osteoarthritis (Bajaj et al. 2001), tension-type headache (Bendtsen et al. 1996), temporomandibular joint pain (Svensson et al. 2001), postmastectomy pain (Gottrup et al. 2000), and chronic low back pain (O'Neill et al. 2007) display pain hypersensitivity after sensory stimulation of healthy tissues, most likely resulting from an alteration of the central processing of sensory input (Curatolo et al. 2004). In this regard, it has been shown that the NWR is a useful tool in the objective assessment of spinal cord hyperexcitability that is present in chronic pain disorders (Banic et al. 2004).

Reliability is essential if a pain test is used for follow up in patients or to investigate the effect of pharmacological interventions. Normative values and test–retest reliability for the NWR threshold (NWR-T) and the electrical pain threshold (EP-T) in pain-free subjects have recently been published (Micalos et al. 2009; Neziri et al. 2010), concluding that both measurements are reliable to be used in experimental pain studies. However, it is necessary to investigate the reliability of the NWR-T and EP-T in patients with chronic pain in order to translate these techniques from laboratory to clinic.

The aims of this study were to determine the test–retest reliability of the NWR-T and EP-T after single and repeated (temporal summation) electrical stimulation in a group of patients with chronic low back pain, and to investigate the association between the NWR-T and the EP-T. The results of these analyses are presented and their relevance is discussed in the context of the role of NWR parameters in clinical research and in the assessment of central hypersensitivity in individual patients.

Materials and methods

Sample size considerations

The sample size for a test–retest reliability study depends on the minimally acceptable level of reliability ρ_0, the target level of reliability ρ (measured as intraclass correlation coefficient), and the number of repetitions of the measurement n, given that the values for type I and type II errors (α and β) are fixed, usually on typical values of 0.05 and 0.20, respectively (Walter et al. 1998). Considering a minimum number of measurements of $n = 2$ (one measurement per session and two different sessions), and with reliability values of $\rho_0 = 0.5$ as the minimally accepted reliability and $\rho = 0.8$ as the expected reliability (Micalos et al. 2009; Rhudy and France 2007), the minimum sample size for this experiment should be 22 subjects. In order to account for unexpected high variability, data from 25 patients were collected.

Patients

The study was approved by the local ethical committee, Inselspital of Bern, Switzerland (KEK Nr. 151/09). Twenty-five patients with low back pain (13 males and 12 females, mean age 51 years, range 23–78 years) participated in the study. They were recruited at the Department of Anesthesiology and Pain Therapy of the University Hospital of Bern, Inselspital. Written informed consent was obtained from all of them prior to participation and the Declaration of Helsinki was respected.

Inclusion criteria were: age of 18–80 years, chronic low back pain of at least 6 months duration, VAS (visual analogue scale) score at the time of testing higher than 3 in a 0–10 range (0: no pain, 10: worst pain imaginable). Exclusion criteria were any additional acute or chronic pain condition, intake of any pain medication for less than 24 h before the investigation, radicular pain, as defined by leg pain associated with an MRI finding of a herniated disc or foraminal stenosis, insufficient knowledge of the German language, pregnancy (as ruled out by pregnancy test in women in reproductive age), and breast feeding.

Experimental procedure

All experiments were performed by the same researcher (A.Y.N.). Each patient participated in three sessions separated by 7.7 ± 2.6 days (mean $\pm$ standard deviation). Each session was carried out in a climate-controlled room at the same time of the day, in order to avoid effects derived from circadian variations (Sandrini et al. 1986). Patients were instructed to maintain their regular daily activities during the testing period, and they were advised to refrain from intake of any pain medication for at least 24 h before the investigation.

Initial setup

The subjects were placed in supine position with a back support, and a pillow was placed under the knee joint of the leg to be stimulated, resulting in a knee flexion of approximately 45°, while the contralateral leg was extended. The stimulation side was selected to match the side most affected by the chronic pain, for both unilateral and bilateral pain situations. EMG and electrical test stimulation electrodes were mounted as described below, and then the subjects were thoroughly familiarized with pain intensity

ratings and electrical stimulation to reduce any effects due to high initial vigilance and/or anxiety (French et al. 2005).

EMG recordings

Activity in the biceps femoris (BF) and vastus lateralis (VL) muscles was measured using surface EMG. Initially, the skin was lightly abraded, and then two surface electrodes (30 × 22 mm, type Neuroline 720, Ambu A/S, Denmark) were placed along the muscle fiber direction over each muscle with an inter-electrode distance of 20 mm. The signal was amplified (up to 20,000 times), filtered (5–500 Hz, second order), sampled (2,000 Hz), and stored (1,000 ms window including 200 ms of pre-stimulation activity).

Reflex detection

NWR detection was performed online by the computer and verified by the researcher carrying out the experiments (A.Y.N.). The criterion used to determine the existence of a NWR was the presence of significant EMG activity (quantified as one or more peaks with amplitude exceeding 20 μV for at least 5 ms) in the 60–180 ms post-stimulation interval of the signals recorded either from BF or from VL muscles, taking into consideration that similar activity was also not present in the pre-stimulation interval.

Thresholds to single electrical stimulation

Electrical stimulation was performed through surface electrodes placed caudal to the lateral malleolus, at the innervation area of the sural nerve. A 25 ms train-of-five square-wave pulses, each lasting 1 ms, were delivered by a computer-controlled constant-current stimulator (Noxitest IES 230, Aalborg, Denmark). The stimulation train was perceived as a single stimulus. The program delivered the pulses at random time intervals (between 8 and 12 s), so that the subject was not aware of when the stimulus was applied. The current intensity was increased from 1 mA in steps of 1 mA until: (1) a reflex was detected (single stimulus NWR-T) and (2) a pain sensation was evoked (single stimulus EP-T) (Neziri et al. 2010). The procedure was repeated three times at 1 min intervals, and the median thresholds were used in order to ensure internal consistency, i.e., to make sure that the estimation of the reflex threshold is stable independent of the chosen definition for reflex detection.

Thresholds to repeated electrical stimulation

The stimulus burst used for single stimulation was repeated five times with a frequency of 2 Hz at constant intensity (Arendt-Nielsen et al. 1994). The current intensity of the five constant stimuli was increased from 1 mA in steps of 1 mA until: (1) two or three consecutive reflexes with increasing amplitude were detected (temporal summation NWR-T) and (2) the subjects felt pain during the last two to three of the five electrical bursts (temporal summation EP-T) (Neziri et al. 2010). As for single electrical stimulation, the procedure was repeated three times at 1 min intervals, and the median thresholds were also used.

Data analysis and statistics

All values are presented as mean ± standard error of the mean (SEM). *P* values smaller than 0.05 were regarded as significant. The NWR-T and EP-T to single and repeated stimulation were compared using repeated measures analysis of variance (RM ANOVA). Threshold (NWR or pain), type of stimulation (single or repeated), and session (first, second or third) were regarded as factors. A multivariate approach was used to bypass the compound symmetry and sphericity assumptions required in a classic RM ANOVA approach (Park et al. 2009).

The between-session reliability (also referred to as stability over time) of the NWR-T and the EP-T to single and repeated electrical stimulation was assessed using multiple methods described below, along with guidelines to help interpreting the results:

- Intraclass correlation coefficient (ICC): it measures the relative homogeneity within sessions in relation to the total observed variation between sessions. For this analysis, a two-way mixed model using absolute agreement was selected, and ICC for single measurements was reported. ICC values above 0.75 are indicative of good reliability (Portney and Watkins 2009). An *F* test with a hypothesized ICC value of 0.5 (representing a fair reliability level) was also carried out in order to determine if the ICC values obtained are significantly different from this level.
- Coefficient of variation (CV): it represents the standard error of measurement expressed as a percentage of the subject's average threshold. The CV can be interpreted as the percentage of deviation from the average threshold below which 68% of the differences between sessions may be expected to lie (Atkinson and Nevill 1998).
- Bland–Altman agreement analysis: it is based on the analysis of the average versus the difference of the thresholds between two given sessions, from which the so-called limits of agreement (LA) can be derived, as the average difference ±1.96 times the standard deviation of the differences. The LA delimit the range within which 95% of the differences between thresholds in two single sessions may be expected to lie. In close relation to this

definition, the coefficient of repeatability (CR) is defined as the value below which 95% of the absolute differences between thresholds in two single sessions may be expected to lie. If there is no systematic bias in the thresholds between sessions, the average difference between thresholds is expected to be zero, and the LA and the CR should be similar (Bland and Altman 1999).

The assessment of test–retest reliability and the comparisons of results from different studies should be done cautiously, depending on the type of parameter used to measure reliability (Atkinson and Nevill 1998). Following their recommendations, the present reliability analysis first includes a full examination of any systematic bias between sessions in the measurements (using an ANOVA test). Then, the ICC is analyzed, given that it has the advantage over other correlation methods (such as Pearson's correlation coefficient), that it can be used when more than one retest is performed. However, these methods depend on the sample heterogeneity (Bland and Altman 1990), and thus are considered measures of relative reliability, since the more homogeneous a population is, the lower the measurement error is needed in order to detect differences between individuals in a population (Atkinson and Nevill 1998). In contrast, measures of absolute reliability (such as standard error of measurement, CV and LA) are not affected by the range of measurements in use. The standard error of measurement and the LA are reported in the same dimension (i.e., units) of the test, whereas the CV is a dimensionless statistic and thus it is useful to compare the reliability among studies using different methodologies (Feltz and Miller 1996).

Finally, the association between the NWR-T and EP-T was measured using the coefficient of determination (r^2, the square of Pearson's correlation coefficient), that represents the percent of the variance in the dependent variable explained by the independent variable. In addition, 95% confidence intervals were calculated, which give a range of values around which the true (population) mean of the dependent variable for given levels of the independent variable can be expected to be located.

Results

Differences across thresholds, types of stimulation, and sessions

Average NWR-T and EP-T to single and repeated stimulation are shown in Fig. 1. RM ANOVA showed that the NWR-T was significantly higher than the EP-T ($p < 0.001$). In addition, single stimulation thresholds were higher than repeated stimulation thresholds ($p < 0.001$). Apart from that, neither a significant difference (bias) was detected between sessions ($p = 0.37$) nor any significant interactions.

Reliability of the NWR-T and EP-T

Detailed ICC and CV values for every combination of sessions are shown in Table 1. Average ICC values were 0.80 for the NWR-T threshold and 0.90 for the EP-T after single stimulation, and the average ICC values after repeated stimulation were 0.76 for the NWR-T and 0.78 for the EP-T. In general, ICC values were above 0.75 and were significantly different from 0.5 (as can be observed in Table 1), thus indicating an overall good reliability. Regarding CV, average values were 17.7% for the NWR-T and 12.0% for the EP-T after single stimulation, and the average CV values after repeated stimulation were 16.9% for the NWR-T and 14.6% for the EP-T.

Bland–Altman plots and the corresponding LA for every combination of session, threshold, and type of stimulation are shown in Figs. 2 and 3. No systematic differences (bias) between sessions were detected in the ANOVA test; therefore, the average differences between thresholds in two given sessions were not significantly different from zero (which can be observed in Figs. 2, 3, dashed line). Average

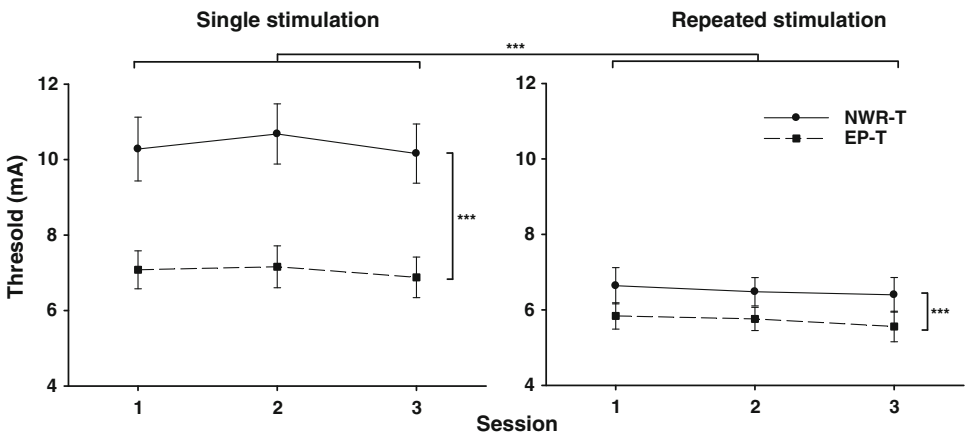

Fig. 1 Average NWR-T and EP-T after single and repeated electrical stimulation (***$p < 0.001$)

Table 1 Detailed analysis for ICC and CV

	NWR threshold (NWR-T)			Electrical pain threshold (EP-T)		
	Sessions 1–2	Sessions 2–3	Sessions 1–3	Sessions 1–2	Sessions 2–3	Sessions 1–3
Intraclass correlation (ICC)						
Single stimulation	0.82**	0.85***	0.71	0.91***	0.94***	0.84**
Repeated stimulation	0.80**	0.84**	0.62	0.81**	0.85**	0.68
Coefficient of variation (CV)						
Single stimulation (%)	16.8	14.4	22.0	11.4	9.4	15.2
Repeated stimulation (%)	14.8	13.4	22.4	12.7	12.5	18.8

** $p < 0.01$; *** $p < 0.001$ in F test for ICC with hypothesized value of 0.5

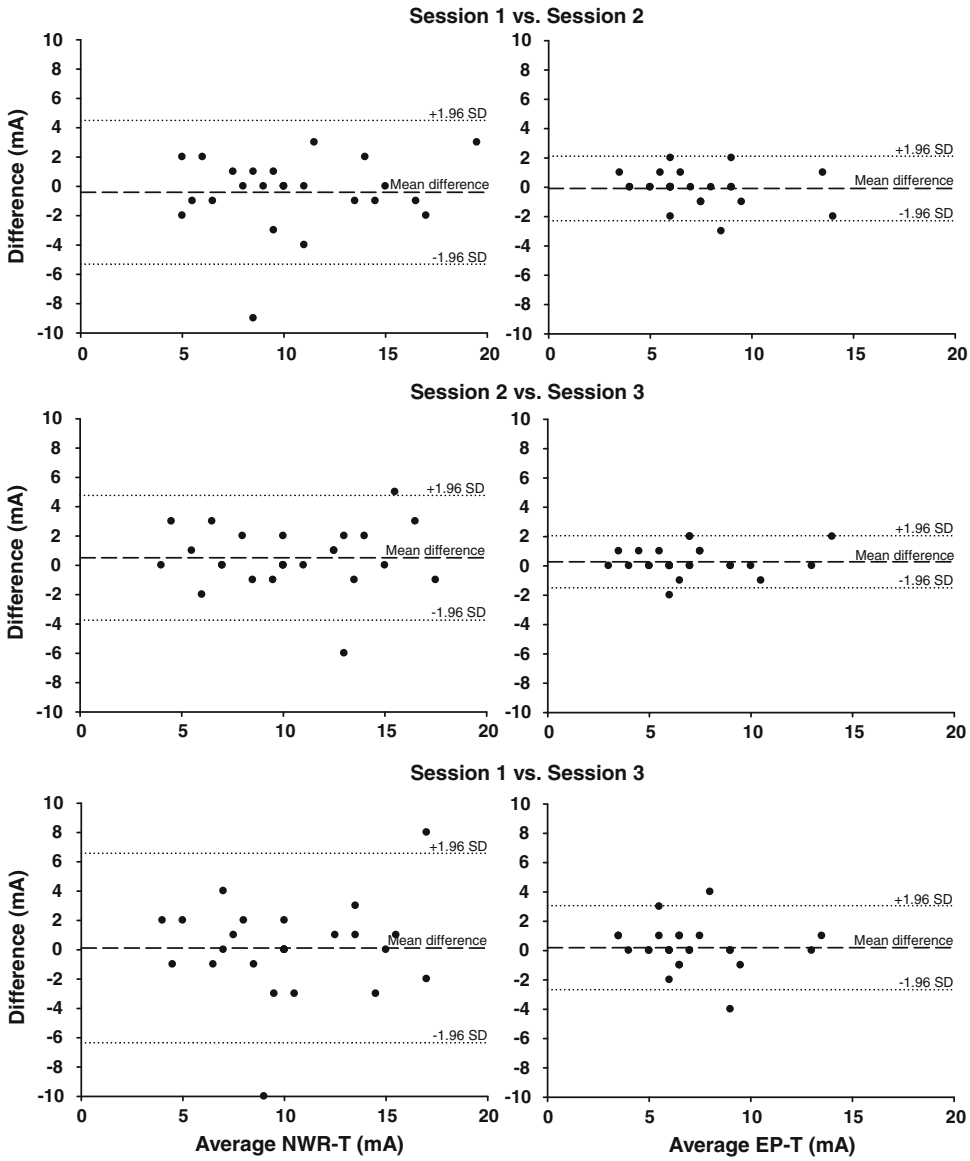

Fig. 2 Bland–Altman plots for the NWR-T and EP-T after single stimulation. *Dashed* Mean difference between the two sessions, *dotted* limits of agreement

CR was 5.0 mA for the NWR-T and 2.3 mA for the EP-T after single stimulation, whereas average CR was 2.9 mA for the NWR-T and 2.3 mA for the EP-T after repeated stimulation.

Association between the NWR-T and EP-T

All correlations were significant at level $p < 0.001$. The degree of association r^2 between the NWR-T and EP-T

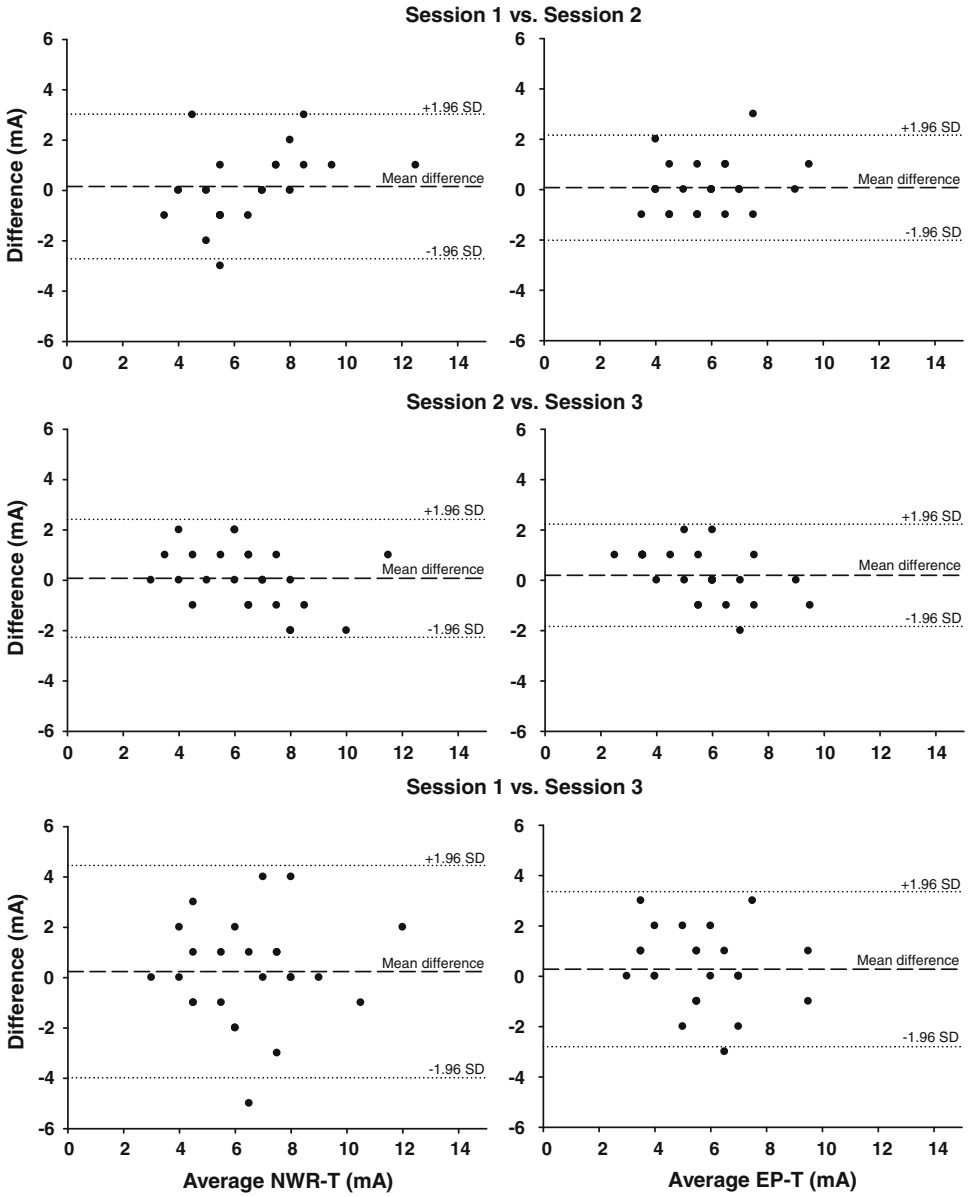

Fig. 3 Bland–Altman plots for the NWR-T and EP-T after repeated stimulation. *Dashed* Mean difference between the two sessions, *dotted* limits of agreement

after single stimulation was 0.53 for session 1, 0.59 for session 2, and 0.56 for session 3. The degree of association r^2 between the NWR-T and EP-T after repeated stimulation was 0.80 for session 1, 0.83 for session 2, and 0.87 for session 3. The regression lines and 95% confidence intervals for all estimations are shown in Fig. 4.

Discussion

In order to determine the reliability of the NWR-T and EP-T in patients with chronic pain, single and repeated (temporal summation) electrical stimulation were applied to 25 patients with chronic low back pain. Three identical sessions were carried on, separated in average by 1 week, where the NWR-T and EP-T after single and repeated stimulation were assessed. RM ANOVA showed that the NWR-T was significantly higher than the EP-T ($p < 0.001$) and that the thresholds obtained after single stimulation were significantly higher than those obtained after repeated stimulation ($p < 0.001$), but no significant differences (bias) were found between sessions. Both NWR-T and EP-T presented good to excellent test–retest reliability. After repeated stimulation, the reliability values were similar for NWR-T and EP-T and generally lower when compared to the results obtained after single stimulation. Threshold reliability was highest when the assessment was done between the second and third sessions and lowest between the first and the last sessions. Finally, the association between the NWR-T and EP-T was better for repeated stimulation than for single stimulation.

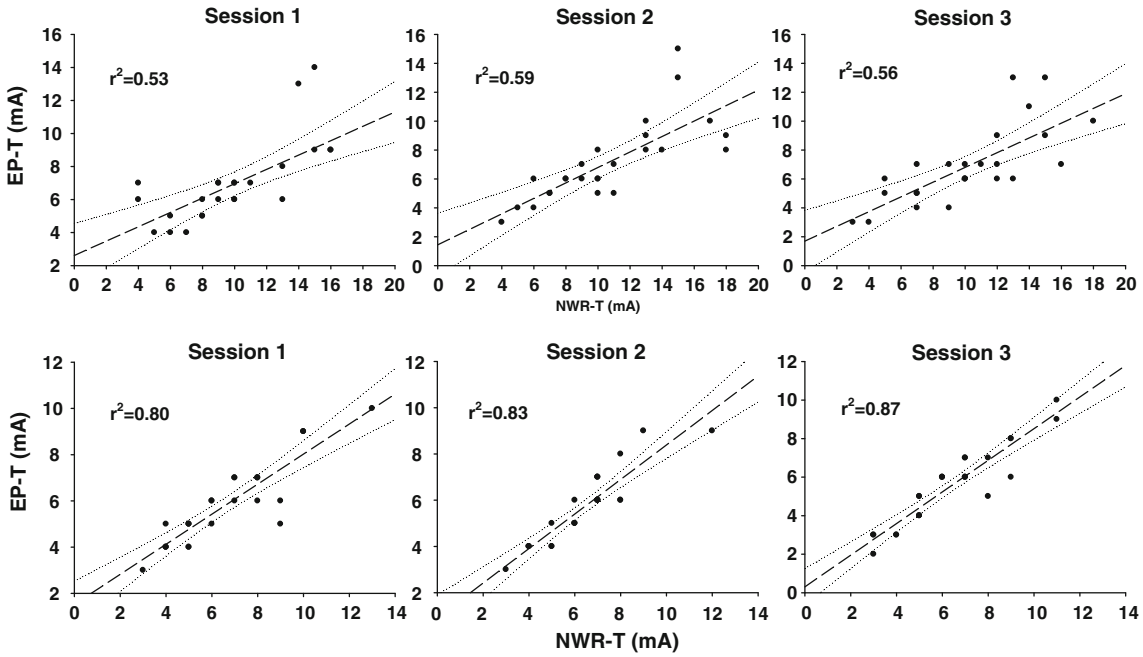

Fig. 4 Association between the NWR-T and EP-T for each session, after single (*top row*) and repeated (*bottom row*) stimulation. *Dashed* Regression line, *dotted* 95% confidence intervals

Relationship between the NWR-T and EP-T

The relationship between the NWR-T and the EP-T has been extensively addressed under several experimental conditions but firm conclusions cannot be drawn. Some studies have shown correlation between the thresholds (Chan and Dallaire 1989; Willer 1977), others have shown that the NWR-T is higher (Campbell et al. 1991; Neziri et al. 2010; Terkelsen et al. 2001) or lower (Bromm and Treede 1980; Micalos et al. 2009) than the EP-T. Moreover, under particular conditions such as hypnosis (Danziger et al. 1998), voluntary muscle contraction (Rossi and Decchi 1994), and spinal cordotomy (Garcia-Larrea et al. 1993), the NWR-T is dissociated from the pain sensation. In addition, supraspinal modulation has been reported to have a differential effect on the NWR and the pain perception (Defrin et al. 2007; French et al. 2005; Terkelsen et al. 2004; Willer et al. 1979).

In the present study, the NWR-T was significantly higher than the EP-T after single and repeated stimulation. From the published literature, it is not possible to conclude that this is always the case; on the contrary, these discrepancies between studies are more likely linked to variations in the experimental procedures, e.g., threshold definitions, electrode positioning or subjective pain intensity scales among others (Micalos et al. 2009). On the other hand, the association between the NWR-T and EP-T was significantly better after repeated stimulation than after single stimulation, in consistency with the results of previous investigations (Arendt-Nielsen et al. 2000). However, the influence of different measurement ranges for single and repeated stimulation cannot be ruled out. The range of intensities for single stimulation is almost twice as large as the range for repeated stimulation, causing lower intra-session variability in the measurements after repeated stimulation, which can affect the estimation of the Pearson's correlation coefficient (Bland and Altman 1986). In any case, the NWR-T and the EP-T are complementary (rather than interchangeable) measurements, and their use will ultimately depend on the type of hypothesis to be tested.

Reliability of the NWR-T and EP-T

The reliability between the NWR-T and EP-T has so far only been addressed in populations of healthy volunteers. Dincklage et al. (2009) reported that the variability between test and retest of the NWR-T after single stimulation, measured as the standard deviation of the differences between measurements, was approximately 4.4 mA when the sessions were approximately 16 weeks apart. Micalos et al. (2009) reported that the reliability analysis of the NWR-T after single stimulation showed in average a CV of 16.9% and an ICC of 0.82, whereas for EP-T, also after single stimulation, the values were in average a CV of 16.1% and an ICC of 0.88, when the sessions were separated approximately by 4 days. Both studies concluded that the NWR-T and EP-T are reliable measurements, and, therefore, can be applied as tools in experimental pain studies. A similar

conclusion was also reached by Lund et al. (2005) in a study involving healthy subjects and pain patients; however, only sensory and pain thresholds to electrocutaneous stimulation were tested, and a custom-designed device with an ordinal scale was used, thus making it difficult to compare these results against similar studies.

In comparison with these studies, the present results render similar reliability values. The EP-T appears to have slightly better reliability than the reflex threshold after single stimulation. A possible explanation lies the fact that the nociceptive input that ultimately elicits the NWR is largely processed in the spinal cord subjected to descending modulation from supraspinal structures (Andersen 2007), whereas report of a painful sensation is subjected to further processing in the brain, that integrates this nociceptive input with additional cognitive and perceptual information (Price 2000, 2002). Thus, several other variables play an important role in pain perception, and some of them (e.g. habituation to electrical stimulation, attention, memory of the ratings of previous stimulations) can affect in such a way that the overall variability of the pain ratings is decreased, resulting in an increase of the repeatability. Interestingly, this effect is not so remarkable for temporal summation, probably due to the fact that repeated stimulation provides a more stable, long-lasting nociceptive input that might allow a more reliable reflex response and a better assessment of the pain sensation.

It should be noted that the differences in reliability among the different tests were, in general, modest. Therefore, it cannot be ruled out that at least some of these differences were the result of chance. In general, the reliability was good to excellent for all tests. Lastly, and although there is no systematic bias in the average NWRT and EP-T between sessions, the reliability is best for the last two sessions and worst when the first and the last sessions are used for the assessment, possibly suggesting a learning effect (Schouenborg 2004) or gradually lower vigilance, despite the initial familiarization with the experimental procedures. Thus, it is expected that the estimated reliability of the NWR-T and EP-T will improve with an increasing number of sessions and a smaller interval of time between sessions (for instance, in crossover studies). Finally, special caution should be taken when follow-up reliability studies are planned involving long periods of time between sessions.

The reliability of the NWR-T and EP-T obtained in studies involving healthy volunteers appear to be comparable to those presented in this study for patients with chronic pain. Moreover, a number of studies have addressed the reliability of other tests that are also used to assess somatosensory function (including cutaneous and deep pain sensitivity), such as the quantitative sensory test or QST (Arendt-Nielsen and Yarnitsky 2009; Rolke et al. 2006). QST test has been widely used to test for sensory differences in a variety of human pain syndromes, such as low back pain (O'Neill et al. 2007), whiplash (Sterling et al. 2003), irritable bowel syndrome (Wilder-Smith et al. 2004), endometriosis (Bajaj et al. 2003), and other pain states (Curatolo et al. 2006). In a recent review, Chong and Cros (2004) presented a meta-analysis of the reproducibility of several QST methods (vibration perception threshold, heat-EP-T, cold perception threshold, and warm perception threshold) in normal subjects as well as in patients suffering from pathological conditions (diabetic with or without neuropathy), concluding that these tests appeared to be sufficiently reproducible during short-term studies (intervals ranging 1–8 weeks). In comparison to the values exhibited by these methods, the reliability of the NWR-T and EP-T reported here is similar or even better, therefore making them suitable for clinical use.

Conclusion

The present results show that the NWR-T and EP-T after single and repeated stimulation are a reliable research tool in experimental and clinical pain that can be applied in the assessment of central hypersensitivity state in individual patients.

Acknowledgments This research was supported by The Danish Research Council for Technology and Production Sciences (FTP), the Swiss National Science Foundation (grant number 3247BO_122358/1), and the Scientific Funds of the University Department of Anesthesiology and Pain Therapy of the University of Bern.

Conflict of interest The authors declare that there is no conflict of interest in the publication of this article.

References

Andersen OK (2007) Studies of the organization of the human nociceptive withdrawal reflex. Focus on sensory convergence and stimulation site dependency. Acta Physiol 189:1–35

Andersen OK, Spaich EG, Madeleine P, Arendt-Nielsen L (2005) Gradual enlargement of human withdrawal reflex receptive fields following repetitive painful stimulation. Brain Res 1042:194–204

Arendt-Nielsen L, Yarnitsky D (2009) Experimental and clinical applications of quantitative sensory testing applied to skin, muscles and viscera. J Pain 10:556–572

Arendt-Nielsen L, Sonnenborg FA, Andersen OK (2000) Facilitation of the withdrawal reflex by repeated transcutaneous electrical stimulation: an experimental study on central integration in humans. Eur J Appl Physiol Occup Physiol 81:165–173

Arendt-Nielsen L, Brennum J, Sindrup S, Bak P (1994) Electrophysiological and psychophysical quantification of temporal summation in the human nociceptive system. Eur J Appl Physiol Occup Physiol 68:266–273

Atkinson G, Nevill AM (1998) Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Med 26:217–238

Bajaj P, Bajaj P, Madsen H, Arendt-Nielsen L (2003) Endometriosis is associated with central sensitization: a psychophysical controlled study. J Pain 4:372–380

Bajaj P, Bajaj P, Graven-Nielsen T, Arendt-Nielsen L (2001) Osteoarthritis and its association with muscle hyperalgesia: an experimental controlled study. Pain 93:107–114

Banic B, Petersen-Felix S, Andersen OK et al (2004) Evidence for spinal cord hypersensitivity in chronic pain after whiplash injury and in fibromyalgia. Pain 107:7–15

Bendtsen L, Jensen R, Olesen J (1996) Decreased pain detection and tolerance thresholds in chronic tension-type headache. Arch Neurol 53:373–376

Bland JM, Altman DG (1999) Measuring agreement in method comparison studies. Stat Methods Med Res 8:135–160

Bland JM, Altman DG (1990) A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement. Comput Biol Med 20:337–340

Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1:307–310

Bromm B, Treede RD (1980) Withdrawal reflex, skin resistance reaction and pain ratings due to electrical stimuli in man. Pain 9:339–354

Campbell IG, Carstens E, Watkins LR (1991) Comparison of human pain sensation and flexion withdrawal evoked by noxious radiant heat. Pain 45:259–268

Chan CWY, Dallaire M (1989) Subjective pain sensation is linearly correlated with the flexion reflex in man. Brain Res 479:145–150

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

Clarke RW, Harris J (2004) The organization of motor responses to noxious stimuli. Brain Res Rev 46:163–172

Curatolo M, Arendt-Nielsen L, Petersen-Felix S (2006) Central hypersensitivity in chronic pain: mechanisms and clinical implications. Phys Med Rehabil Clin N Am 17(2):287–302

Curatolo M, Arendt-Nielsen L, Petersen-Felix S (2004) Evidence, mechanisms, and clinical implications of central hypersensitivity in chronic pain after whiplash injury. Clin J Pain 20:469–476

Curatolo M, Petersen-Felix S, Arendt-Nielsen L, Giani C, Zbinden AM, Radanov BP (2001) Central hypersensitivity in chronic pain after whiplash injury. Clin J Pain 17:306–315

Danziger N, Fournier E, Bouhassira D et al (1998) Different strategies of modulation can be operative during hypnotic analgesia: a neurophysiological study. Pain 75:85–92

Defrin R, Peleg S, Weingarden H, Heruti R, Urca G (2007) Differential effect of supraspinal modulation on the nociceptive withdrawal reflex and pain sensation. Clin Neurophysiol 118:427–437

Dincklage Fv, Hackbarth M, Schneider M, Baars JH, Rehberg B (2009) Introduction of a continual RIII reflex threshold tracking algorithm. Brain Res 1260:24–29

Feltz CJ, Miller GE (1996) An asymptotic test for the equality of coefficients of variation from k populations. Stat Med 15:647–658

French DJ, France CR, France JL, Arnott LF (2005) The influence of acute anxiety on assessment of nociceptive flexion reflex thresholds in healthy young adults. Pain 114:358–363

Garcia-Larrea L, Charles N, Sindou M, Mauguiere F (1993) Flexion reflexes following anterolateral cordotomy in man: dissociation between pain sensation and nociceptive reflex RIII. Pain 55:139–149

Gottrup H, Andersen J, Arendt-Nielsen L, Jensen TS (2000) Psychophysical examination in patients with post-mastectomy pain. Pain 87:275–284

Graven-Nielsen T, Sörensen J, Henriksson KG, Bengtsson M, Arendt-Nielsen L (1999) Central hyperexcitability in fibromyalgia. J Musculoskelet Pain 7:261–271

Hagbarth KE (1960) Spinal withdrawal reflexes in the human lower limbs. J Neurol Neurosurg Psychiatry 23:222–227

Lund I, Lundeberg T, Kowalski J, Sandberg L, Budh CN, Svensson E (2005) Evaluation of variations in sensory and pain threshold assessments by electrocutaneous stimulation. Physiother Theory Pract 21:81–92

Micalos PS, Drinkwater EJ, Cannon J, Arendt-Nielsen L, Marino FE (2009) Reliability of the nociceptive flexor reflex (RIII) threshold and association with pain threshold. Eur J Appl Physiol 105:55–62

Neziri AY, Andersen OK, Petersen-Felix S et al (2010) The nociceptive withdrawal reflex: normative values of thresholds and reflex receptive fields. Eur J Pain 14:134–141

O'Neill S, Manniche C, Graven-Nielsen T, Arendt-Nielsen L (2007) Generalized deep-tissue hyperalgesia in patients with chronic low-back pain. Eur J Pain 11:415–420

Park E, Cho M, Ki C (2009) Correct use of repeated measures analysis of variance. Kor J Lab Med 29:1–9

Portney LG, Watkins MP (2009) Foundations of clinical research: applications to practice. Pearson/Prentice Hall, Upper Saddle River, NJ

Price DD (2002) Central neural mechanisms that interrelate sensory and affective dimensions of pain. Mol Interv 2:392–403, 339

Price DD (2000) Psychological and neural mechanisms of the affective dimension of pain. Science 288:1769–1772

Rhudy JL, France CR (2007) Defining the nociceptive flexion reflex (NFR) threshold in human participants: a comparison of different scoring criteria. Pain 128:244–253

Rolke R, Magerl W, Campbell KA et al (2006) Quantitative sensory testing: a comprehensive protocol for clinical trials. Eur J Pain 10:77–88

Rossi A, Decchi B (1994) Flexibility of lower limb reflex responses to painful cutaneous stimulation in standing humans: evidence of load-dependent modulation. J Physiol 481:521–532

Sandrini G, Serrao M, Rossi P, Romaniello A, Cruccu G, Willer JC (2005) The lower limb flexion reflex in humans. Prog Neurobiol 77:353–395

Sandrini G, Alfonsi E, Bono G (1986) Circadian variations of human flexion reflex. Pain 25:403–410

Schouenborg J (2004) Learning in sensorimotor circuits. Curr Opin Neurobiol 14:693–697

Schouenborg J, Kalliomäki J (1990) Functional organization of the nociceptive withdrawal reflexes. Exp Brain Res 83:67–78

Serrao M, Rossi P, Sandrini G et al (2004) Effects of diffuse noxious inhibitory controls on temporal summation of the RIII reflex in humans. Pain 112:353–360

Sherrington CS (1910) Flexion-reflex of the limb, crossed extension-reflex, and reflex stepping and standing. J Physiol 40:28–121

Sterling M, Jull G, Vicenzino B, Kenardy J (2003) Sensory hypersensitivity occurs soon after whiplash injury and is associated with poor recovery. Pain 104:509–517

Svensson P, List T, Hector G (2001) Analysis of stimulus-evoked pain in patients with myofascial temporomandibular pain disorders. Pain 92:399–409

Terkelsen AJ, Andersen OK, Mølgaard H, Hansen J, Jensen TS (2004) Mental stress inhibits pain perception and heart rate variability but not a nociceptive withdrawal reflex. Acta Physiol Scand 180:405–414

Terkelsen AJ, Andersen OK, Hansen PO, Jensen TS (2001) Effects of heterotopic- and segmental counter-stimulation on the nociceptive withdrawal reflex in humans. Acta Physiol Scand 172:211–217

Tørring J, Pedersen E, Klemar B (1981) Standardisation of the electrical elicitation of the human flexor reflex. J Neurol Neurosurg Psychiatry 44:129–132

Walter SD, Eliasziw M, Donner A (1998) Sample size and optimal designs for reliability studies. Stat Med 17:101–110

Wilder-Smith CH, Schindler D, Lovblad K, Redmond SM, Nirkko A (2004) Brain functional magnetic resonance imaging of rectal pain and activation of endogenous inhibitory mechanisms in irritable bowel syndrome patient subgroups and healthy controls. Gut 53:1595–1601

Willer JC, Boureau F, Albe-Fessard D (1979) Supraspinal influences on nociceptive flexion reflex and pain sensation in man. Brain Res 179:61–68

Willer JC (1977) Comparative study of perceived pain and nociceptive flexion reflex in man. Pain 3:69–80