

Systematic Review

A Systematic Review of Devices and Techniques that Objectively Measure Patients' Pain

Sjors H. Wagemakers, MD¹, Joanne M. van der Velden, MD, PhD¹, A. Sophie Gerlich, MD¹, Alinde W. Hindriks-Keegstra, MD², Jacqueline F.M. van Dijk, PhD², and Joost J.C. Verhoeff, MD, PhD¹

From: ¹Department of Radiation Oncology, University Medical Center Utrecht, The Netherlands;

²Department of Anesthesiology, University Medical Center Utrecht, The Netherlands

Address Correspondence:
Joost J.C. Verhoeff, MD, PhD
Department of Radiotherapy
University Medical Center
Utrecht

Heidelberglaan 100
3584 CX Utrecht

E-mail:
J.J.C.Verhoeff-10@umcutrecht.nl

Disclaimer: There was no external funding in the preparation of this manuscript. Conflict of interest: Each author certifies that he or she, or a member of his or her immediate family, has no commercial association (i.e., consultancies, stock ownership, equity interest, patent/licensing arrangements, etc.) that might pose a conflict of interest in connection with the submitted manuscript.

Manuscript received: 12-29-2017

Revised manuscript received:
05-10-2018

Accepted for publication:
06-08-2018

Free full manuscript:
www.painphysicianjournal.com

Background: Assessment of pain is important in daily clinical practice and as an endpoint in clinical studies. Because pain perception is highly subjective, pain measurement is complex. Self-rating pain scales are currently of great importance but have limitations. They depend on many more factors than pain, which could lead to an incorrect assessment of therapies or clinical studies. Therefore, there is need for valid, reliable, safe, and low-cost methods to determine and quantify patients' pain more objectively.

Objective: To provide an overview of devices and techniques that can be used to administer a pain stimulus with similar intensity as the endogenous pain experienced by the patient, in order to quantify and subsequently follow patients' pain more objectively.

Study Design: In this systematic review, articles from PubMed, EMBASE, Cochrane library and Scopus were reviewed for eligibility.

Methods: Studies that described a device or technique that could be used to induce a variable, controlled, and measurable pain stimulus were included. Studies that made correlations with established pain scales or those who compared outcomes in multiple tests were selected to assess validity and reliability.

Results: A total of 1,308 manuscripts were initially retrieved. After independent screening by a team of 4 reviewers, 19 studies were eventually included describing 15 different devices or techniques. These devices could be divided into groups based on stimulus administration: electrical, external pressure (probe) and miscellaneous pain stimulators. Electrical stimulators were found to be tested extensively and proven to be both valid and reliable.

Limitations: To correlate new techniques with older methods such as the Numeric Rating Scale (NRS) and Visual Analogue Scale (VAS) for which an improvement is desired, is debatable. To (partially) address this problem, the reliability is added as an additional primary outcome to assess which device works best. Further limitations include the heterogeneity of studies found in both the types of pain measured as in outcome measures presented. In addition, it is important to note that part of the devices described cannot directly be used for clinical practice due to products that have cease to exist or the description of solely techniques rather than testing ready-to-use devices.

Conclusion: Several devices and techniques compared pain intensity experienced by patients with an external pain stimulus that potentially could be used as a new objective pain measurement tool. Given the results of our review, electrical stimulators that have been tested extensively with high validity, reliability, and feasibility would be recommended for use for clinical and research purposes. Moreover, normalization of pain intensity scores for current perception is important. Pain intensity normalization leads to higher correlations with established pain scales and possibly to increased inter-patient reliability.

Registration number: Registered in the PROSPERO database (PROSPERO 2016:CRD42016041974)

Key words: Systematic review, objective pain measurement, pain scales; devices, techniques, validity, reliability, safety, feasibility

Pain Physician 2019; 22:1-13

Pain is a very common physical sensation. In the United States, 11.2% of the population experiences pain every day (1). Assessment of pain is complex. Patients' pain perception is highly subjective and based on personal experience (2). Because pain measurement is often used in daily clinical practice and as primary or secondary endpoints in clinical studies, an accurate and precise measure of pain is of great importance (3–5).

Self-rating instruments are of great importance to help physicians assess patients' health state (6–8). Simple and one-dimensional measures like the Numeric Rating Scale (NRS) and the Visual Analog Scale (VAS) are available. While it is possible to quantify acute pain using one-dimensional measurements, for chronic pain it is advised to take the multidimensional nature of pain into account (2). Multidimensional measures have been developed, including the extensive McGill Pain Questionnaire (MPQ) and the Brief Pain Inventory (BPI). The MPQ assesses pain in more dimensions and is based on descriptive words that patients can choose to express their pain best (9).

The VAS and NRS are highly correlated and probably the most frequently used pain scales for measurement of pain in the daily clinic (6). However, these simple pain scales show some limitations in assessment of pain intensity. They are highly subjective and depend on many more factors than pain, such as mood (10–12). This could lead to an imprecise assessment of effectiveness of therapies in daily practice or clinical studies. In multiple pain measurements, patients score their actual pain intensity using previously experienced pain as a reference. However, there is some evidence that the recall of previously experienced pain is only moderately accurate (13). Moreover, these pain experiences differ for each individual, which influences the frame of reference for each person. As a result, it is shown that differences in pain experiences result in different pain tolerances per person (14). Pain scores using the VAS and NRS are presumably affected by the patients predetermined levels of pain. Because participants themselves should indicate pain score, it is not possible to blind participants for the outcome of the VAS or NRS. Therefore they can (subconsciously) contribute to the success or failure of a therapy, by indicating lower pain scores which overestimate the effect of treatment (15). From these perspectives, it is difficult to quantify the degree of pain in a precise manner. Therefore, there is a need for a valid, reliable, safe, and low-cost method to determine and quantify patients' pain more objectively.

The aim of this review is to present an overview of devices and techniques that can be used to administer a pain stimulus with similar intensity as the endogenous pain experienced by the patient, in order to quantify and subsequently follow patients' pain more objectively.

METHODS

Search strategy

The review protocol was registered in the PROSPERO database (PROSPERO 2016:CRD42016041974) and applies a systematic approach according to the PRISMA guideline (16). A search was conducted in July 2016 and updated in July 2017, of PubMed, EMBASE, Cochrane library and Scopus databases. The search strategy included search terms and their synonyms for stimulation of pain in combination with pain (intensity) measurement or dolorimetry and correlations with other established pain scales. Pain stimulating devices that were used for therapy, as well as animal studies were excluded in the search string (Appendix A). Additional articles were retrieved by cross-referencing the included articles.

Study Selection

All studies describing a device or technique used to induce a variable, controlled, and measurable pain stimulus were included. We selected English or Dutch studies that used other established pain scales such as the VAS, NRS and MPQ as a control group and those studies that compare outcomes in multiple tests. Both healthy participants and patients suffering from pain were allowed. While patients were able to compare a painful stimulus in terms of intensity with their own ongoing pain, healthy participants were unable to match this intensity. Therefore, these people were asked to score the perceived pain to an increasing stimulus to examine the correlation and reproducibility of devices and techniques.

Pain stimulating devices that were used for therapy were excluded. Animal studies were excluded. Other methods to quantify patients' pain, for instance using heart rate, fMRI or the concentration of particular markers, were excluded as well since these are indirect manifestations of pain. No date restriction was applied and all study designs were allowed.

Screening and Data Extraction

Studies were independently screened by title and abstract by a team of 4 reviewers to select relevant

studies for inclusion, followed by full text screening. Conflicts in study inclusion were discussed by the study team.

Primary endpoints were validity and reliability or reproducibility of the studied device or technique. Data about either or both of the primary outcomes were extracted including correlation with established pain scales and test-retest intraclass correlation coefficient (ICC). Secondary outcomes included safety and feasibility.

We assessed quality of publications and the risk of bias using a self-made checklist consisting of items we considered important for quality assessment, including description of populations, experiment protocols, and safety aspects. A final overall judgement for each study was made by the study team. Due to the heterogeneity of data, the results of the systematic literature search were combined in a hybrid method with narrative syntheses and analyses (17).

RESULTS

Eligibility and Inclusion of Studies

By combining the results of the 4 databases, a total of 1308 articles were identified. After removal of duplicates, 994 records were screened for title and abstract (Fig. 1). The most common reasons for exclusion included interventions such as application of a constant stimulus, a stimulator serving as treatment, or use of methods that are indirect manifestations of pain. After full-text screening of 126 articles, 13 studies were included. Five additional studies were included after cross-referencing on specific brand names of devices yielded by the initial search. As search terms for specific names of devices were not included in the initial search string, it cannot be ruled out that some other studies have been missed. One study was added after the search was updated in July 2017.

A total of 19 studies were included, describing 15 different devices and techniques. Most studies were prospective case control studies, with sample size ranging from 12 to 242 participants. Included studies showed heterogeneity in various areas. In some studies, measurements were performed on healthy volunteers whereas in other studies measurements were performed on groups of patients suffering from, for example, acute oral pain, subacute whiplash pain or dysmenorrhea. Moreover, the location of pain appliance varied between different extremities, as well as the usage of various control scales (VAS, VRS, NRS, and MPQ) in the included studies.

Quality Assessment

Five of the 19 studies were judged as high-quality studies as they almost completely described all the factors considered to affect study quality (Table 1). Except for one study (18), no other studies explicitly described safety issues. The use of perception thresholds (to determine a sensation baseline) were only applied in a minority of studies (17,19-22). Population and experiment procedure were sufficiently described in almost every included study. Eleven studies used random stimulus intensity to prevent participants' anticipatory reactions. To test the validity of each study across a broad range of pain intensities, the ranges of applied pain scores were assessed. Most studies used a VAS/NRS range of 6 or more.

Techniques of Applying Pain Stimuli

The included studies were divided into 3 groups based on mode of stimulus administration: electrically based pain stimulators (n = 9), external pressure (probe) based pain stimulators (n = 4) and a group of miscellaneous techniques consisting of dilatation, brushing, sound and shockwave induced pain stimulators (n = 5).

Electrical Stimulators

Table 2 and Fig. 2 summarizes the results of 9 studies describing 6 different electrical stimulators.

PainMatcher

In total, 4 publications described pain stimulation using the PainMatcher (Cefar Medical AB, Sweden) for electrical stimulation of sensory nerve fibers (18-20,23). The device determines so-called perceptual matching. Participants hold the device between the thumb and index finger and during an increasing stimulus, they are asked to release fingers at either the sensory threshold, pain threshold, or at a level that matches the ongoing pain intensity. To increase the intensity, the pulse duration slowly increases in 60 steps from 0 to 450 μ s. Releasing the PainMatcher stops electrical stimulation and the current intensity at that time is automatically stored.

Different types of pain were studied in combination with the PainMatcher including acute oral pain (18), subacute whiplash pain (19), chronic musculoskeletal pain (23) and pregnant women's pain during labor (20). The intensity of acute oral pain, which was normalized via pain threshold, shows good and significant correlation to the VAS pain scale (18). Also, for

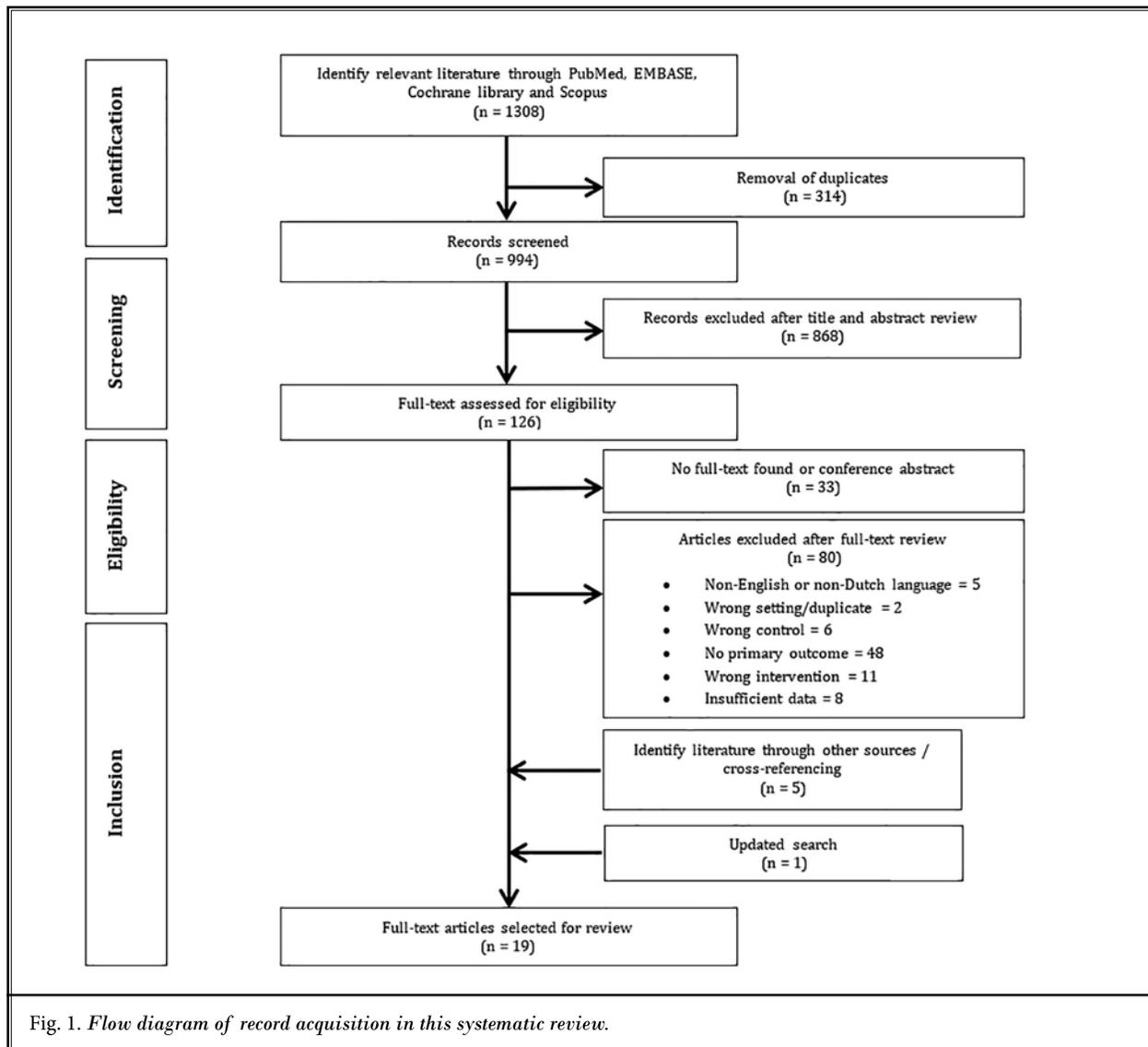


Fig. 1. Flow diagram of record acquisition in this systematic review.

the test-retest reproducibility, significant correlations were described of $r = 0.93$ and $r = 0.97$. Although the PainMatcher showed high reproducibility, it is also associated with unpleasantness because of the (additional) pain stimulus (19).

In a chronic pain group consisting of disabled patients with musculoskeletal pain (23), a very weak correlation between the VAS and PainMatcher $r = 0.17$ was found when pooling data from all patients, indicating a large individual spread. However, in an additional analysis of individual data-sets a high correlation was found with a median of $r = 0.64$.

PainVision

Two articles used the PainVision® (PS-2100, Nipro Co., Japan) for electrical stimulation (21,24). Similar techniques as described for the PainMatcher were used, although the stimulation place is located on the medial forearm. In the PainVision, the current varies from 0–150 μA . Using the PainVision, first a current perception threshold (CPT) is defined. Subsequently, the pain equivalent current (PEC) is defined as the current with equal pain intensity as the ongoing pain. Quantitative pain degree (QPD) was defined as $(\text{PEC} - \text{CPT})/\text{CPT}$, leading to a dimensionless and normalized pain score.

Objective Measurement of Patients' Pain

Table 1. Factors that are considered to affect study applicability are included in this table. Quality analysis of reviewed articles with green rated 'good', orange 'reasonable' and red as 'bad'.

	Reference	Description of population	Sensation baseline described	Safety issues described	Range of scores [VAS/NRS]	Reproducibility of stimulus	Experiment procedure	Random intensity increase	Judged quality
Electrical	Alstergren et al (18)	Good	Good	Bad	[0-10]	Good	Reasonable	Good	High
	Kall et al (19)	Good	Bad	Good	[1-9]	Good	Good	Good	High
	Persson et al (23)	Good	Good	Reasonable	[1-9]	Reasonable	Good	Bad	Medium
	Bergh et al (20)	Good	Bad	Reasonable	[0-10]	Reasonable	Good	Good	Medium
	Kim et al (21)	Good	Good	Reasonable	[0-9]	Good	Good	Bad	High
	Ono et al (24)	Good	Good	Bad	[0-8]	Reasonable	Reasonable	Bad	Low
	van der Heide et al (22)	Reasonable	Bad	Reasonable	[1-5]	Bad	Good	Good	Low
	Oliveira et al (25)	Good	Good	Reasonable	[0-8]	Bad	Good	Good	High
	Steenbergen et al (26)	Reasonable	Bad	Reasonable	[0-10]	Reasonable	Good	Good	Medium
Mechanical	Naganawa et al (27)	Good	Bad	Reasonable	[1-7]	Good	Good	Good	High
	Goddard et al (28)	Good	Bad	Reasonable	[0-6]	Good	Good	Bad	Medium
	O'Neill et al (29)	Good	Bad	Reasonable	[0-10]	Good	Good	Good	High
	Graven-Nielsen et al (31)	Good	Bad	Reasonable	[2-5]	Good	Good	Bad	Low
Miscellaneous	Bajaj et al (32)	Good	Bad	Reasonable	[1-6]	Good	Good	Bad	Medium
	Tu et al (35)	Good	Bad	Reasonable	Bad	Reasonable	Good	Good	Low
	du Jardin et al (37)	Good	Bad	Reasonable	[2-4]	Good	Good	Bad	Low
	Bromage et al (38)	Reasonable	Bad	Bad	[0-6]	Bad	Reasonable	Good	Low
	Lara-Munoz et al (39)	Good	Bad	Bad	[1-9]	Good	Good	Good	Medium
	Vaegter et al (40)	Good	Bad	Reasonable	Bad	Good	Good	Bad	Low

When a detailed description had been given about the population (including number of participants, age, health condition, in-, exclusion criteria) the study was rated 'good'. When only 2 of these conditions were reported, the study was rated 'reasonable'. When the intensity of the stimulus was normalized against the baseline, this was determined to be 'good'. Safety aspects of the techniques discussed were rated 'good' when the study explained how safety issues were addressed (or when referred to a study that proved the safety of the techniques), and 'reasonable' when the study had been at least approved by a local (ethical) committee. Because it is important that a technique has been validated over a broad range of pain scores, the (mean) range on the VAS or NRS scales that were applied were assessed. A predetermined chosen range of at least 6 points were rated 'good', range between 3 and 6 distinctive pain points 'reasonable' and lower than 3 points 'bad'. Reproducibility was rated 'good' when experiments were included in the study to confirm reproducibility and 'reasonable' when only referred to a study that proved reproducibility. Description of the experiment protocol was rated 'good' when detailed information about the setting, technique, examiners and instruction to the participants had been given, and 'reasonable' when at least 2 of these conditions were described. Finally, studies that administered a stimulus with random intensity increase to prevent participants' anticipatory reactions were rated 'good'. Eventually, a final overall judgment has been made. Studies that involved at most one 'bad' and one 'reasonable' aspect were judged as high-quality studies, studies including 3 or more 'bad' aspects or 2 'bad' aspects in combination with 2 or more 'reasonable' aspects were judged as low quality and studies in between were judged as medium quality studies.

Table 2. Overview of the results of the electrically stimulators (with rS = Spearman's rank correlation coefficient, rP = Pearson correlation coefficient, RV = relative rank variance, RP = relative position, ICC = intraclass correlation coefficient, * = significant).

Reference	Type of stimulus	Quantity (Units)	Location of stimulus	Corrected for pain baseline	No. participants	Participants characteristics	Control pain scale	Correlation	Reliability
Alstergren et al (18)	Electrical current (PainMatcher®)	Pulse width (ms)	Thumb and index finger	Pain threshold	n = 28	Acute oral pain	VAS	$rS = 0.63$ ($P = 0.004^*$)	Test-retest Correlation (successive assessments) $r = 0.93 - 0.97$ ($p < 0.001^*$) $RV = 0.033 - 0.078$ $RP = -0.001 - 0.002$
Kall et al (19)	Electrical (PainMatcher®)	Pulse width (ms)	Thumb and index finger	-	n = 47	Subacute whiplash-associated disorder	VAS	$rS = 0.46$ ($P < 0.01^*$)	Test-retest $RV = 0.01$ $RP = -0.003$
Persson et al (23)	Electrical (PainMatcher®)	Pulse width (ms)	Thumb and index finger	Perception threshold	n = 31	Chronic musculoskeletal pain	VAS	$rS = 0.17$ ($P < 0.0001^*$) Individual data-sets (n = 20): $rS = 0.64$	-
Bergh et al (20)	Electrical (PainMatcher®)	Pulse width (ms)	Thumb and index finger	-	n = 128	Healthy (pregnant) woman	VAS	$rS = 0.13$ ($P < 0.05^*$)	-
Kim et al (21)	Electrical (PainVision®)	Current (μA)	Medial forearm	Perception threshold	n = 25	Low back pain	VAS	$rS = 0.240 - 0.289$ ($P = 0.248 - 0.161$)	Test-retest $ICC: 0.967$ ($P < 0.001^*$)
Ono et al (24)	Electrical (PainVision®)	Current (μA)	Forearm	Perception threshold	n = 42	Low back pain	MPQ	$rS = 0.619 - 0.628$ ($P = 0.001^*$)	-
van der Heide et al (22)	Electrical	Current (mA) and number of pulses	Forearm or fingertip	-	n = 41	Healthy	NRS	$rS = 0.50$ ($P = 0.005^*$) Changes after 4 weeks: $rS = 0.444$ ($P = 0.003^*$) $rS = 0.48$ ($P < 0.01^*$) Changes after 4 weeks: $rS = 0.411$ ($P = 0.01^*$) Significant linear relationship ($P < 0.0001^*$)	-
Oliveira et al (25)	Electrical	Current (μA)	Lower leg	Pain threshold	n = 10	Healthy	VAS	$rP = 0.99$ ($P < 0.05^*$)	-
Steenbergen et al (26)	Electrical	Current (mA)	Lower arm	-	n = 10	Healthy	VAS	-	Test-retest Disc electrode ICC: 0.80 [0.58-0.91] Needle electrodes ICC: 0.84 [0.65-0.92]

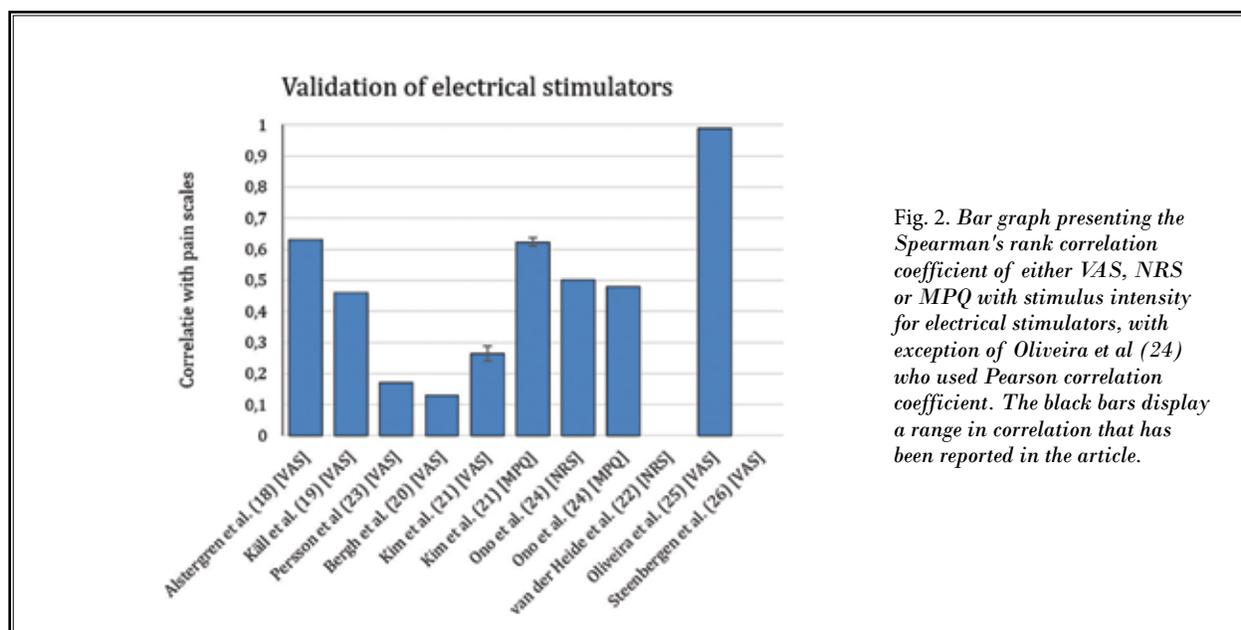


Fig. 2. Bar graph presenting the Spearman's rank correlation coefficient of either VAS, NRS or MPQ with stimulus intensity for electrical stimulators, with exception of Oliveira et al (24) who used Pearson correlation coefficient. The black bars display a range in correlation that has been reported in the article.

Both studies performed tests on patients suffering from lower back pain (21,24). A moderate to strong correlation between the established pain score (MPQ) and the quantitative pain degree outcome of the PainVision (QPD) was found in these studies ($r \approx 0.62-0.63$ and $r = 0.48$). A significant correlation between VAS and QPD; however, it was not found (21). The authors attribute this lack of correlation to the one-dimensional VAS, which is known to assess chronic pain less accurately (2). However, an agreement in changes in MPQ relative to QPD was found after 4 weeks of conservative treatment. Test-retest reliability tested for the PainVision was excellent with a ICC of 0.967 in 3 identical tests within 5 minutes.

Other Electric Stimulators

Three other electrical stimulation methods were used, applying a pain stimulus on the forearm or fingertip, lower leg and lower arm (22,25,26). All 3 methods varied the stimulus intensity by changing the electrical current. One study also tried to change the number of pulses in order to increase stimulus intensity (22). A linear relationship has been found between NRS and both the stimulus amplitudes and number of pulses, which indicates that both methods can be used to change stimulus intensity. An excellent correlation between the relative pain score and the participants reported VAS of $r = 0.99$ was shown in Oliveira et al (25). Tests of the

reproducibility of electric stimulation relative to VAS with 5 pulses tests showed an excellent ICC of both disc electrode 0.80 (95% confidence interval [CI] 0.58-0.91) and needle electrode 0.84 (95% CI 0.65-0.92).

Pressure Stimulators

Table 3 summarizes the findings of 4 manuscripts describing mechanical stimulators.

External Probe or Clamp

Similar techniques were described in 3 articles, using an external probe or clamp on the gingival mucosa, masseter muscles or thumbnail to evoke a pain sensation (27–29). Only the pressure clamp showed a high and significant correlation of $r = 0.73$ among the amount of pressure (kg) and the patients' experienced pain measured using VAS (29). All 3 studies demonstrated good test-retest reproducibility (Table 3). Three repeated tests gave ICC of 0.76 and 0.99 (27) and 0.811 and 0.844 (28). The pressure clamp also showed decent test-retest values, with mean differences of only 1.54 and 1.84 mm on a scale of 100 mm (VAS).

Cuff Pressure Algometer

Although the study scored low at the quality criteria of this review, Lim et al described a pressure technique applying a computer-controlled cuff pressure algometer (30). It applies a tourniquet cuff around

both the arm and leg for a limited time to induce pain. Although no correlation with other pain scales have been described, it was shown that the duration of tourniquet application corresponds with an increasing pain score (30). Via an electronic continuous VAS measurement, healthy participants had to assess their pain intensity until intolerable pain. After 10 repeated identical stimuli with a pressure equivalent to the pain tolerance, temporal summation was demonstrated, indicating a significant higher pain score in the last tests compared to the first tests (31).

Miscellaneous Stimulators

Table 3 and Fig. 3 summarizes the findings of 6 stimulators that apply other forms of stimulation.

Dilators

Dilatation of body parts can be used in clinical studies, for example dilatation of the esophagus, duodenum and rectum to generate a measurable pain stimulus (32-34). Tu et al (35) applied pressure on the vagina using a thin thimble with a pressure sensor in which the examiner provides a varying pressure. Moderate correlations were shown between the pressure of the examiner and the reported VAS ($r = 0.61$). Besides a thimble, one can exert pressure using a balloon which is inflated in the uterus (32). The cross-sectional area (CSA) of the balloon in the uterus correlated to the VAS score with a correlation of 0.9. When registering 3 repeated distensions of the esophagus, no significant differences in sensory and pain threshold were found indicating a good test-retest reproducibility for balloon dilators (36). Furthermore, a high correlation of the CSA and balloon volume with VAS was found. This is in contrast to the lower correlation with the more obvious balloon pressure, which showed an exponential increase in VAS scores.

Automated Brushing Devices

Another external pressure stimulus is the fully automated brushing device (37). Twelve healthy participants with induced allodynia after capsaicin injection in the forearm were asked to rate their dynamic mechanical allodynia on a VAS scale after each set of strokes. Different angles and brushing speeds were used. Excellent reproducibility (ICC = 0.84-0.97) was shown in 3 identical tests, concluding that such a device is capable to quantify allodynic reactions. Moreover, automatic brushing showed superior results compared to manual

brushing (with ICC reproducibility of 0.72-0.91). No correlations with other pain scales were calculated.

Ultrasound

A study by Bromage et al (38), demonstrated the use of a lithotripter as a controlled pain delivery device concentrated on the abdomen. Although the lithotripter originally was designed to crush kidney stones and gallstones, a good correlation of 0.84 was reported between voltage and VAS scores.

Audible Sound

Also based on pressure is applying a pain stimulus to the ear using audible sound (39). A pure tone of 1000 Hz had been applied for 3 seconds in 5 different intensities (dB). Good to very good correlations were described among sound intensity and either VAS, verbal rating scale (VRS) and NRS, respectively scoring 0.82, 0.80 and 0.74. Significant pain was generated, with VAS scores up to 9 on a scale of 10. Little attention was paid to the safety aspects in this study. The intra-observer variability of this method was assessed in 3 successive tests which represent a moderate to good agreement.

Thermal

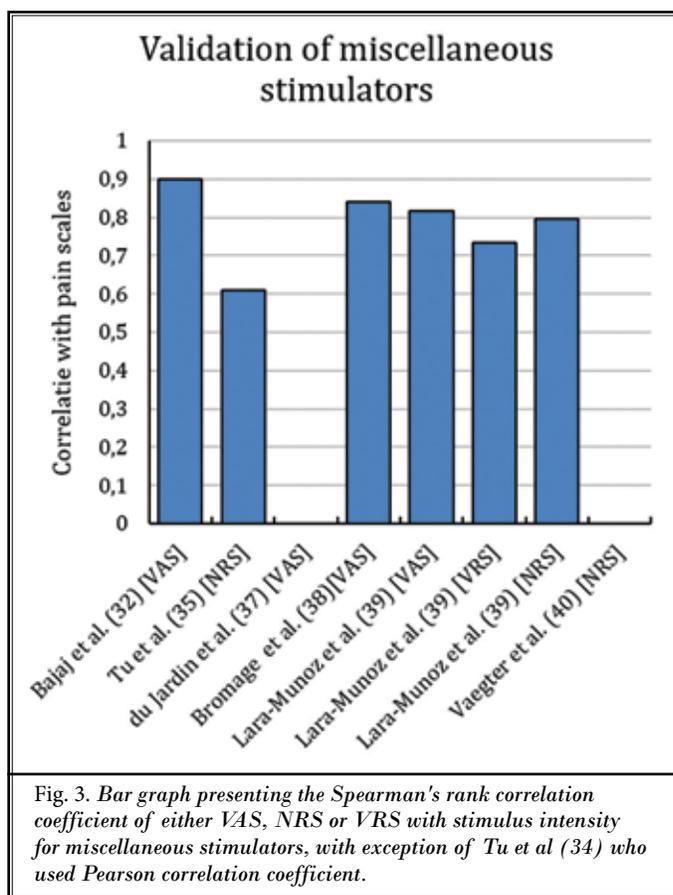
As a thermal stimulator, a computer controlled surface thermode (MSA Thermal Stimulator, SENSELab, Somedic Sales AB, Hörby, Sweden) was utilized to induce a pain stimulus (40). Temperature of the thermode on the foot started at 32°C and increased by 1.0°C per second with a maximum of 50°C. Participants had to define at which temperature the first sensation of pain was observed. The thermal stimulus was repeated 3 times with ICC of 0.54 and 0.54.

DISCUSSION

Pain is thought to be a subjective physical sensation. Objective measurements of pain in patients are complex. Alternatives for currently used simple pain scales are needed due to subjectivity and the dependence on many more factors than just pain alone. In this systematic review, 19 studies with 15 different techniques or devices have been described, which can be used to induce variable, controlled and measurable pain stimulus. The goal of this review is to outline which methods are most valid (i.e., regarding to existing pain scales), most reliable (i.e., in multiple successive tests), most safe, and most feasible, to give an overview of the best "bedside tests" that measure patient's pain more objectively.

Table 3. Overview of the results of the mechanical and miscellaneous stimulators (with $r_s =$ Spearman's rank correlation coefficient, $r_p =$ Pearson correlation coefficient, $r_t =$ Linear correlation (not otherwise specified), ICC = intraclass correlation coefficient, * = significant).

Reference	Type of stimulus	Quantity (Units)	Location of stimulus	Corrected for pain baseline	No. participants	Participants characteristics	Control pain scale	Correlation	Reliability
Mechanical									
Naganawa et al (27)	Pressure (probe)	Pressure force (kg), diameters (mm) and shapes	Gingival mucosa	-	n = 16	Healthy	NRS	-	Test-retest ICC: 0.76 to 0.99
Goddard et al (28)	Pressure (probe: Wagner algometer*)	Pressure force (pounds)	Masseter muscles	-	n = 18	Healthy	VAS	-	Test-retest ICC: 0.811 - 0.844 Gender did not affect reproducibility
O'Neill et al (29)	Pressure (clamp)	Pressure (kg)	Thumb nail	-	n = 20	Healthy	VAS	$r_s = 0.73$ ($P < 0.0001^*$)	Test-retest Mean VAS difference: Healthy participants: 1.84 (SD = 13.81) Hospital patients 1.54 (SD = 24.45)
Graven-Nielsen et al (31)	Pressure (tourniquet cuff)	Pressure (kPa)	Arm and leg	-	n = 136	Healthy	VAS	-	Test-retest Temporal summation confirmed using Newman-Keuls test ($P < 0.036^*$)
Miscellaneous									
Bajaj et al (32)	Dilatation	Cross-sectional area (mm ²)	Uterus	-	n = 9	Healthy	VAS	$r_p = 0.9$ ($P < 0.001^*$)	Test-retest No significant differences (Friedman $P > 0.1$ and $P > 0.6$)
Tu et al (35)	Dilatation	Pressure (kg/cm ²)	Vagina	-	n = 19	Healthy	NRS	$r_s = 0.61$ ($P < 0.05^*$)	-
du Jardin et al (37)	Pressure (brushing device)	Angle (°) and speed (mm/s)	Forearm	-	n = 12	Injection of capsaicin	VAS	-	Test-retest Automatic brushing ICC: 0.84-0.97 Manual brushing ICC: 0.72-0.91
Bromage et al (38)	Pressure (shock wave)	Voltage (kV)	Abdomen	-	n = 5	Healthy	VAS	$r_t = 0.84^*$	-
Lara-Munoz et al (39)	Pressure (sound: pure 1000Hz tone)	Intensity (dB)	Ears	-	n = 30	Healthy	VAS	$r_s = 0.818$ ($P < 0.00001^*$)	Test-retest Weighted Kappa VAS: 0.70 (86.9%) Weighted Kappa VRS: 0.65 (81.8%) Weighted Kappa NRS: 0.59 (86.4%)
Vaegter et al (40)	Heat pain threshold	Temperature (°C)	Dorsum of foot	-	n = 20	Healthy	NRS	$r_s = 0.735$ ($P < 0.00001^*$)	Test-retest ICC: 0.54



The techniques causing a painful stimulus can be roughly divided into 3 categories: mechanical, electrical and thermal stimulators. Of the 19 studies found, 9 studies described an electrical stimulation and 10 studies described some form of mechanical stimulation (including the described mechanical and miscellaneous group). Of these electrical stimulators, different devices that are already further developed and proved in multiple studies were shown to be both valid and reliable (18-21,23,24). Although thermal stimulation has widely been examined in the literature (41-43), data on the correlation with established pain scales is lacking. As a result, only 1 thermal study was included in this review, focusing on the reproducibility with an ICC of only 0.54. Although this company is bankrupt, competitors are available using thermal stimulation with Peltier thermodes (44).

From the nature of this review, it seems illogical to compare a new technique with older methods such as the NRS and VAS, as we are looking for an improvement on these methods. Yet we deliberately chose this approach. When no correlation could be found relative to a new technique, we can assume that these methods are not valid. For example, if a patient in-

dicates a significantly higher VAS, this experience should somehow correspond to a higher pain score on the new pain assessment tool. On the other hand, when a perfect correlation is demonstrated (r to 1.00), the new technique would not have any added value since both pain tools give the same outcomes. In that case, the easy-to-use VAS would be preferred in daily clinical practice. To address this issue, we also took reliability into account as an additional primary outcome. Often reliability is expressed in terms of reproducibility or test-retest abilities. Good reproducibility suggests that pain scores are not just random, but that they show consistent value associated with the particular participants.

It is possible to correlate new high-tech pain measurements with other techniques to assess pain more objectively. Heart rate variability, fMRI, EEG and EMG are generally considered to be more accurate and objective measures for pain quantification (25,45,46). Although it is unlikely that these techniques will be used in daily practice, validation of easy-to-use devices could be performed with the help of these techniques.

It is important to take other factors of the pain experience into account, such as life events, family and work environment, or even seasonal influences. These factors might play a role as potential confounders in pain perception (47,48). If these potential confounders were registered in a cohort study, it might enable correction for these factors to validate new pain measurements.

Interestingly, a pain threshold can serve as a baseline sensitivity measure to normalize pain scores. When testing the validity of the PainMatcher, a large difference was observed between correlations (with VAS) of direct pain intensity scores and correlations (with VAS) of pain intensity scores normalized for pain threshold (18). The correlation increased significantly after normalization for pain threshold from $r = 0.18$ to $r = 0.63$. First, this normalization is important in order to ensure intra-patient reliability. The location of an electrode could differ in successive tests. Different stimulus locations could affect measurements by the influence of different sensory nerve distribution and skin conditions (21), where dryness or presence of hair on the skin could alter the impedance and,

as a result, affect the electrical stimulation. Second, it may improve the inter-patient reliability. Pain is an individual perception and highly subjective. However, normalizing the pain with a baseline pain threshold could partly eliminate the difference in pain sensation for each individual, improving inter-patient reliability. For most studies, a relative increase or decrease in pain is the primary outcome, for which a high inter-patient reliability is not necessary. In cross-sectional studies however, when pain scores are determined once, high inter-patient reliability offers promising possibilities.

It is remarkable that for patients suffering from chronic pain, no correlation was found between VAS and electrical stimulus intensity (23); whereas, in patients with acute pain, a clear correlation was found between stimulus intensity and VAS (18). Though it is possible to quantify chronic pain using one-dimensional measurements, the complex multidimensional nature associated with chronic pain is not taken into account. Therefore, for chronic pain assessment it is advised to use the extensive McGill Pain Questionnaire, for example (2). This tool quantifies pain in more dimensions and has a good correlation with the MPQ in patients suffering chronic pain (21).

Only one-third of the studies made use of a sensation baseline (18,21,23-25). It is important for new pain instruments to be tested in a broad range of pain scores (Table 1). Most studies did well, especially those that tested with a pain scale from 0 up to 10. The brushing device however was only tested for VAS ranging from 2-4, and it is unclear how this device responds to VAS pain intensity above 4. Stimulus range also plays an important role. For example, the audible sound stimulus could only vary in 5 different intensities, limiting a new pain score tool to only 5 possible values.

Safety and feasibility was found to be scarcely reported. Only 1 of the included studies provided information about these outcomes. Our selection of articles based on criteria that included at least the details of primary outcomes, resulted in limited information about our secondary outcomes. We can assume that safety issues were considered before research was performed on humans, especially when the approval of a local (ethical) committee was given. However, this interpretation is open to debate, and more detailed information is needed. Little was reported about practicality; however, one can assume that the use of a lithotripter or of invasive methods

such as a dilator with the uterus, are less feasible to perform compared to smaller and simpler handheld electrical devices or a test using sound intensity.

Limitations

There are several limitations to this systematic review. First, the heterogeneity of the included studies has to be taken into account. The intention of this review was to include all types of pain administration as well as the measurement of all sorts of pain. By omitting further restrictions to modes of administration or measurements, devices and methods found were tough to compare. Moreover, the assessment of validity and reliability in the studies did not all correspond. Hence pooling data from the studies to perform a meta-analysis was not possible, and therefore, results were presented as narrative syntheses and analyses. Because this review was designed and executed in a systematic way, it can be classified best as a systematic narrative review.

A second limitation is that production of the Pain-Matcher has ceased. This is a constraint because this device is not applicable for further clinical practice. Furthermore, some techniques mentioned in this review are not directly applicable for routinely clinical practice, because no ready-to-use devices are available. Devices based on these methods will first have to be designed, which does not guarantee the same outcomes as presented in this review.

CONCLUSION

Patients' pain perception is highly subjective and based on personal experience. This presents a problem for both daily clinical practice and research purposes, where effects of interventions need to be evaluated. This review presents a heterogeneous group of devices and techniques which can be used as objective pain measurement methods. It is shown that especially electrical stimulators are best suited, with good validation to existing pain scales and high reliability in multiple successive tests. This review shows the importance of the normalization of pain intensity scores for current perception, as pain intensity normalization leads to higher correlations with established pain scales and possibly to increased inter-patient reliability.

Acknowledgments

We would like to thank Paulien H. Wiersma from the Utrecht University Library for her contribution to the search strategy and Dr. Simon J. Woodings for reviewing this manuscript for English language.

REFERENCES

1. NIH analysis shows Americans are in pain, National Institutes of Health: www.nih.gov/news-events/news-releases/nih-analysis-shows-americans-are-pain.
2. Bhardwaj P, Yadav R. Measuring pain in clinical trials: Pain scales, endpoints, and challenges. *Int J Clin Exp Physiol* 2015; 2:151–156.
3. Higginson IJ, Simon ST, Benalia H, Downing J, Daveson BA, Harding R, Bausewein C. Which questions of two commonly used multidimensional palliative care patient reported outcome measures are most useful? Results from the European and African PRISMA survey. *BMJ Support Palliat Care* 2012; 2:36–42.
4. Meek R, Egerton-Warburton D, Mee MJ, Braitberg G. Measurement and monitoring of nausea severity in emergency department patients: A comparison of scales and exploration of treatment efficacy outcome measures. *Acad Emerg Med* 2015; 22:685–693.
5. Triano JJ, McGregor M, Cramer GD, Emde DL. A comparison of outcome measures for use with back pain patients: Results of a feasibility study. *J Manipulative Physiol Ther* 1993; 16:67–73.
6. Ahlers SJGM, van Gulik L, van der Veen AM, van Dongen HPA, Bruins P, Belitser S V, de Boer A, Tibboel D, Knibbe CAJ. Comparison of different pain scoring systems in critically ill patients in a general. *ICU Crit Care* 2008; 12:R15.
7. Sari O, Temiz C, Golcur M, Aydogan U, Tanoglu A, Ezgu MC, Tehli O. Pain perception differences between patients and physicians: A pain severity study in patients with low back pain. *Turk Neurosurg* 2015; 25:461–468.
8. Suarez-Almazor ME, Conner-Spady B, Kendall CJ, Russell AS, Skeith K. Lack of congruence in the ratings of patients' health status by patients and their physicians. *Med Decis Making* 2001; 21:113–121.
9. Melzack R. The McGill Pain Questionnaire: Major properties and scoring methods *Pain* 1975; 1:277–299.
10. Taenzer P, Melzack R, Jeans ME. Influence of psychological factors on post-operative pain, mood and analgesic requirements. *Pain* 1986; 24:331–342.
11. Katz J, Melzack R. Measurement of Pain. *Surg Clin North Am* 1999; 79:231–252.
12. Tang NKY, Salkovskis PM, Hodges A, Wright KJ, Hanna M, Hester J. Effects of mood on pain responses and pain tolerance: An experimental study in chronic back pain patients. *Pain* 2008; 138:392–401.
13. Erskine A, Morley S, Pearce S. Memory for pain: A review. *Pain* 1990; 41:255–265.
14. Rollman GB, Abdel-Shaheed J, Gillespie JM, Jones KS. Does past pain influence current pain: Biological and psychosocial models of sex differences. *Eur J Pain* 2004; 8:427–433.
15. Lundeberg T, Lund I, Dahlin L, Borg E, Gustafsson C, Sandin L, Rosén A, Kowalski J, Eriksson SV. Reliability and responsiveness of three different pain assessments. *J Rehabil Med* 2001; 33:279–283.
16. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Ann Intern Med* 2009; 151:264–269.
17. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, Britten N, Roen K, Duffy S. Guidance on the conduct of narrative synthesis in systematic reviews: A Product from ESRC Methods Program. Version 2006; 1:b92. http://www.lancaster.ac.uk/shm/research/nssr/research/dissemination/publications/NS_Synthesis_Guidance_v1.pdf
18. Alstergren P, Forstrom J. Acute oral pain intensity and pain threshold assessed by intensity matching to pain induced by electrical stimuli. *J Orofac Pain* 2003; 17:151–159.
19. Kall LB, Kowalski J, Stener-Victorin E. Assessing pain perception using the Painmatcher in patients with whiplash-associated disorders. *J Rehabil Med* 2008; 40:171–177.
20. Bergh IHE, Stener-Victorin E, Wallin G, Martensson L. Comparison of the Pain-Matcher and the Visual Analogue Scale for assessment of labour pain following administered pain relief treatment. *Midwifery* 2011; 27:e134–139.
21. Kim J, Lee KS, Kong SW, Kim T, Kim MJ, Park S-B, Lee KH. Correlations between electrically quantified pain degree, subjectively assessed Visual Analogue Scale, and the McGill Pain Questionnaire: A pilot study. *Ann Rehabil Med* 2014; 38:665–672.
22. van der Heide EM, Buitenweg JR, Marani E, Rutten WLC. Single pulse and pulse train modulation of cutaneous electrical stimulation: A comparison of methods. *J Clin Neurophysiol* 2009; 26:54–60.
23. Persson AL, Westermark S, Merrick D, Sjolund BH. Validity of electrical stimulus magnitude matching in chronic pain. *J Rehabil Med* 2009; 41:898–903.
24. Ono Y, Ohtori S, Orita S, Yamauchi K, Aoki Y, Miyagi M, Suzuki M, Kubota G, Sakuma Y, Oikawa Y, Sainoh T, Sato J, Nakamura J, Shiga Y, Eguchi Y, Abe K, Fujimoto K, Kanamoto H, Takahashi K, Inage K. PainVision apparatus is effective for assessing low back pain after fusion surgery. *Chiba Med J* 2014; 90:31–35.
25. Oliveira IM, Machado ARP, Pereira AA, Andrade AO. Estimated features from surface EMG of the lower limb correlate with the subjective sensation of pain. *Psychol Neurosci* 2014; 7:355–361.
26. Steenbergen P, Buitenweg JR, Trojan J, van der Heide EM, van den Heuvel T, Flor H, Veltink PH. A system for inducing concurrent tactile and nociceptive sensations at the same site using electrocutaneous stimulation. *Behav Res Methods* 2012; 44:924–933.
27. Naganawa T, Iida T, Baad-Hansen L, Ando T, Svensson P. Application of a new palpometer for intraoral mechanical pain sensitivity assessment. *J Orofac Pain* 2013; 27:336–342.
28. Goddard G, Karibe H, McNeill C. Reproducibility of visual analog scale (VAS) pain scores to mechanical pressure. *Cranio* 2004; 22:250–256.
29. O'Neill S, Graven-Nielsen T, Manniche C, Arendt-Nielsen L. Reliability and validity of a simple and clinically applicable pain stimulus: Sustained mechanical pressure with a spring-clamp. *Chiropr Man Ther* 2014; 22.
30. Lim E, Shukla L, Barker A, Trotter DJ. Randomized blinded control trial into tourniquet tolerance in awake volunteers. *ANZ J Surg* 2015; 85:636–638.
31. Graven-Nielsen T, Vaegter HB, Finocchietti S, Handberg G, Arendt-Nielsen L. Assessment of musculoskeletal pain sensitivity and temporal summation by cuff pressure algometry: A reliability study. *Pain* 2015; 156:2193–2202.
32. Bajaj P, Drewes AM, Gregersen H, Petersen P, Madsen H, Arendt-Nielsen L. Controlled dilatation of the uterine cervix--an experimental visceral pain model. *Pain* 2002; 99:433–442.
33. Gao C, Arendt-Nielsen L, Liu W, Petersen

- en P, Drewes AM, Gregersen H. Sensory and biomechanical responses to ramp-controlled distension of the human duodenum. *Am J Physiol Gastrointest Liver Physiol* 2003; 284:G461-G471.
34. Grinsvall C, Tornblom H, Tack J, Van Oudenhove L, Simren M. Psychological factors selectively upregulate rectal pain perception in hypersensitive patients with irritable bowel syndrome. *Neurogastroenterol Motil* 2015; 27:1772-1782.
35. Tu FF, Fitzgerald CM, Kuiken T, Farrell T, Norman Harden R. Vaginal pressure-pain thresholds: Initial validation and reliability assessment in healthy women. *Clin J Pain* 2008; 24:45-50.
36. Drewes AM, Pedersen J, Liu W, Arendt-Nielsen L, Gregersen H. Controlled mechanical distension of the human oesophagus: Sensory and biomechanical findings. *Scand J Gastroenterol* 2003; 38:27-35.
37. du Jardin KG, Gregersen LS, Rosland T, Uggerhoj KH, Petersen LJ, Arendt-Nielsen L, Gazerani P. Assessment of pain response in capsaicin-induced dynamic mechanical allodynia using a novel and fully automated brushing device. *Pain Res Manag* 2013; 18:6-10.
38. Bromage PR, Husain I, el-Faqih S, Bon-su AK, Kadiwal GH, Seraj P. Critique of the Dornier HM3 lithotripter as a clinical algometer. *Pain* 1990; 40:255-265.
39. Lara-Munoz C, De Leon SP, Feinstein AR, Puente A, Wells CK. Comparison of three rating scales for measuring subjective phenomena in clinical research. I. Use of experimentally controlled auditory stimuli. *Arch Med Res* 2004; 35:43-48.
40. Vaegter HB, Hoeger Bement M, Madsen AB, Fridriksson J, Dasa M, Graven-Nielsen T. Exercise increases pressure pain tolerance but not pressure and heat pain thresholds in healthy young men. *Eur J Pain* 2017; 21:73-81.
41. Defrin R, Pope G, Davis KD. Interactions between spatial summation, 2-point discrimination and habituation of heat pain. *Eur J Pain* 2008; 12:900-909.
42. Lautenbacher S, Galfe G, Karlbauer G, Moltner A, Strian F. Effects of chronic back pain on the perception of experimental heat pain. *Percept Mot Skills* 1990; 71:1283-1292.
43. Lautenbacher S, Moltner A, Strain F. Psychophysical features of the transition from pure heat perception to heat pain perception. *Percept Psychophys* 1992; 52:685-690.
44. Pavlaković G, Klinke I, Pavlaković H, Züchner K, Zapf A, Bachmann CG, Graf BM, Crozier TA. Effect of thermal mode application pressure on thermal threshold detection. *Muscle Nerve* 2008; 38:1498-1505.
45. Jess G, Pogatzki-Zahn EM, Zahn PK, Meyer-Frießem CH, Meyer-Friessem CH. Monitoring heart rate variability to assess experimentally induced pain using the analgesia nociception index. *Eur J Anaesthesiol* 2016; 33:118-125.
46. Roberts K, Papadaki A, Goncalves C, Tighe M, Atherton D, Shenoy R, McRobbie D, Anand P. Contact heat evoked potentials using simultaneous eeg and fmri and their correlation with evoked pain. *BMC Anesthesiol* 2008; 8:8.
47. Feuerstein M, Sult S, Houle M. Environmental stressors and chronic low back pain: life events, family and work environment. *Pain* 1985; 22:295-307.
48. Moldofsky H. Chronobiological influences on fibromyalgia syndrome: Theoretical and therapeutic implications. *Baillière's Clin Rheumatol* 1994; 8:801-810.

