Review

Quantitative sensory testing in painful osteoarthritis: a systematic review and meta-analysis

A.K. Suokas*, D.A. Walsh†, D.F. McWilliams†, L. Condon†, B. Moreton†, V. Wylde‡, L. Arendt-Nielsen§, W. Zhang†

† Arthritis Research UK Pain Centre, University of Nottingham, UK
‡ Musculoskeletal Research Unit, University of Bristol, UK
§ Center for Sensory-Motor Interaction, Aalborg University, Denmark

Article info

Article history:
Received 23 January 2012
Accepted 21 June 2012

Keywords:
Quantitative sensory testing
Osteoarthritis
Pain mechanism
Prediction
Pain phenotyping
Sensitisation

Summary

Objective: To systematically review the use of Quantitative sensory testing (QST) in pain characterisation (phenotyping) in Osteoarthritis (OA).

Methods: Six bibliographic databases (Medline, Embase, Amed, Cinahl, PubMed, Web of Science) were searched to identify studies published before May 2011. Data were extracted based on the primary site of OA, QST modalities, outcome measures and test sites. Standardised mean difference (SMD) and 95% confidence intervals (CIs) were calculated if possible. Publication bias was determined using funnel plot and Egger’s test. Heterogeneity was examined using Cochran Q test and I² statistic. Random effects model was used to pool the results.

Results: Of 41 studies (2281 participants) included, 23 were case control studies, 15 case only studies, two randomised controlled trials, and one uncontrolled trial. The majority of studies examined pressure pain with smaller numbers using electrical and/or thermal stimuli. QST was more often applied to the affected joint than distal and remote sites. Of 20 studies comparing people with OA and healthy controls, seven provided sufficient information for meta-analysis. Compared with controls, people with OA had lower pressure pain thresholds (PPTs) both at the affected joint (SMD = 1.24, 95%CI 1.54, 0.93) and at remote sites (SMD = 0.88, 95%CI −1.11, −0.65).

Conclusion: QST of PPTs demonstrated good ability to differentiate between people with OA and healthy controls. Lower PPTs in people with OA in affected sites may suggest peripheral, and in remote sites central, sensitisation. PPT measurement merits further evaluation as a tool for phenotyping OA pain.

© 2012 Osteoarthritis Research Society International. Published by Elsevier Ltd. All rights reserved.

Introduction

Osteoarthritis (OA) is the most common form of arthritis and a major cause of chronic musculoskeletal pain and disability, worldwide. Although joint damage and inflammation are widely recognised as major contributing factors, the pathophysiology of OA pain remains relatively poorly understood. Multiple mechanisms ranging from local to neurogenic can contribute to OA pain severity, and factors extrinsic to the joint, such as periartricular and central sensitisation, may be important in the maintenance of pain. Sensitisation, in turn, may trigger a transition from acute to chronic pain and influence responses to treatment. Studies suggest, for example, that radiographic evidence does not always match the individual’s report of pain and 10–20% of people with knee OA still have persistent severe knee pain after total knee replacement. The complexity of OA pain means that treatments targeting one specific mechanism may have low efficacy if offered to people whose pain is largely mediated by other mechanisms. Treatment guidelines therefore recommend that the optimal management of OA should be tailored according to patient characterstics including pain phenotyping.

Quantitative sensory testing (QST) is one approach that has been used to evaluate mechanism- (as opposed to symptoms-) based phenotyping of OA pain. QST involves assessing somatosensory evoked responses to noxious or innocuous stimuli using controlled mechanical, chemical, electrical, and/or thermal test modalities. The examiner systematically applies a stimulus to an anatomical test site until the study participant indicates
sensation or pain. Participants’ responses to external stimuli can be assessed at the affected anatomical site to identify localised, or at a distal or remote site to identify regional or widespread, somatosensory changes17. Localised pain on the affected joint is associated with peripheral sensitisation, whereas pain at a site adjacent to or at a distance from the site of origin suggests a combination of peripheral and central sensitisation18. In QST, methods have been developed for investigating complex pain mechanisms such as peripheral and central sensitisation, descending pain control, and referred pain19. QST has also been used to explore differences between diseased and control populations20–26, to investigate pain mechanisms7,22,24,27, and to predict interventions22–24 or measure responses to interventions24,28,35–37. However, published studies have provided variable results, leading to uncertainty as to the extent that pain thresholds and sensory detection thresholds in people with OA may differ from those in healthy people.

The aims of this systematic review and meta-analysis were to examine: [1] the use of QST in OA (modalities, outcome measures and anatomical test sites); [2] the reliability of QST; [3] the ability of QST to differentiate people with OA from healthy controls, and [4] differences between anatomical test sites.

Methods

Systematic literature search

Six bibliographic databases (Medline 1948–, Embase 1980–, Amed 1985–, Cinahl 1981–, PubMed 1950– and Web of Science 1970–) were searched to identify studies published in peer-reviewed journals before May 2011. A systematic search strategy was developed in Medline (see Appendix 1) and replicated as closely as possible in the other bibliographic databases. Additional studies were identified by searching the references of the included articles. Reports were downloaded into Endnote X4.

Inclusion/exclusion criteria

Primary studies of any type of research design involving people with OA were eligible when they: 1) applied at least one of the following QST modalities: chemical, electrical, mechanical or thermal stimulus; 2) measured perception of noxious or innocuous stimuli applied to skin, muscle or joint; 3) used a testing protocol to control for stimulus properties: modality, anatomical site, intensity, duration, and sequence; 4) reported the findings in a peer-reviewed academic journal. Each abstract was assessed by AS for potential relevance. Full text was retrieved for articles that appeared relevant for further analysis. There was no language restriction.

Full papers that met our criteria were included. A meta-analysis comparing people with OA and healthy controls was undertaken if possible. Additional data were obtained from the authors, if they were not reported fully in the papers27,38,39.

Data-extraction

Standard information was extracted from all eligible studies using a single form suitable for all study designs. The information included the following: study design, setting; sample selection; affected joint; diagnostic criteria, demographic data; ethnicity; weight; body mass index (BMI); pain severity on visual analogue scale (VAS) prior to QST; stimulus protocol; QST modalities and outcome measures; the anatomical site of QST; and the mean values and estimates of random variability of pain and/or sensory detection threshold measurement using QST. The studies were classified into randomised controlled trials (RCTs) and observational studies (cohort, case control, case only). Setting was defined as hospital or community setting according to the source of the population, or the source of the control population if it was a case control study. English language data were extracted and coded by one reviewer (AS), and a sub-set of key variables was validated by four co-investigators (BM, DAW, DMcW and LC). Discrepancies and disagreements were resolved by a fifth co-investigator (WZ). Data-extraction in other languages required help from native speakers with knowledge of rheumatology or neuroimaging research (see acknowledgements).

Quality and content assessment

Quality of the studies included in the meta-analysis was examined using a modified version of the criteria devised by Downs and Black40, which is generic to both RCTs and observational studies. Each criterion was scored as 1 if present and 0 if absent or unclear, and the maximum score was 12 (Appendix 2).

Data analysis

QST modalities, outcome measures and test sites

Data were categorised according to study design, setting, and QST modalities including 1) electrical, 2) chemical, 3) mechanical (sub-groups included pressure, punctate/brush, and vibratory) and 4) thermal stimulus. Anatomical QST sites were coded according to their location in relation to the joint(s) reported as affected by OA in each study, and organised into four categories: affected joint, distal, remote, and other sites. Sites coded as ‘affected joint’ were on, or in close proximity to, the primary site of OA in each study. ‘Distal’ sites were below the affected joint and ‘remote’ sites above or contra-lateral to the affected joint. For studies that did not distinguish between uni- and bilateral OA, a contralateral site was coded as ‘remote’ if it was described as pain free. Sites were coded as ‘other’ if the study measured an average reading from two symmetrical sites (e.g., both hands) or an average reading across affected, distal and remote sites, or if the location was unclear. The data were coded by one reviewer (AS) and validated by four co-investigators (BM, DAW, DMcW, LC).

Reliability of QST, and the ability to differentiate between groups

Each study was screened for intraclass correlation coefficients (ICCs) describing the test–retest repeatability of QST. The reliability of the measurement across test occasions was rated excellent if the ICC > 0.75, adequate if 0.40–0.74, and poor if <0.4041. The ability of QST to detect somatosensory differences between people with OA and healthy controls was assessed using standardised mean differences (SMDs). SMD was calculated to estimate the magnitude of the difference (Cohen’s d) between people with OA and healthy controls. By convention, a SMD or d = 0.2 is considered to be small, d = 0.5 moderate and d = 0.8 large in size42. This analysis could only be performed for one outcome measure, pressure pain thresholds (PPTs), due to the small number of case control studies spread across many QST modalities (see Table 1). PPTs were extracted in and, where necessary, converted to kg/cm². The SMDs and 95% confidence intervals (CIs) were presented using a forest plot. Publication bias was examined using a funnel plot and an Egger’s test. The Q test and I² statistics were calculated to measure the degree of heterogeneity between studies. The Q test suggests a significant level of heterogeneity if statistically significant (P < 0.05) whereas the I² value (0–100%) indicates a percentage of the heterogeneity across studies that is not due to chance43,44.

QST results from different anatomical test sites

QST results from different anatomical test sites were assessed by examining forest plots for SMDs using the methods described above. Statistical pooling was undertaken as appropriate according
The mean age of participants in the studies was 62 years; in case control studies, 859 in case only studies, and 157 in the other studies. The total number of participants was 1265 in all studies.

Table 1
Study characteristics

<table>
<thead>
<tr>
<th></th>
<th>Case control</th>
<th>Case only</th>
<th>Other</th>
<th>All studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. studies</td>
<td>23</td>
<td>15</td>
<td>3</td>
<td>41</td>
</tr>
<tr>
<td>No. subjects</td>
<td>1265</td>
<td>859</td>
<td>157</td>
<td>2281</td>
</tr>
<tr>
<td>Mean age</td>
<td>60</td>
<td>64</td>
<td>60</td>
<td>62</td>
</tr>
<tr>
<td>Female %</td>
<td>64</td>
<td>64</td>
<td>45</td>
<td>62</td>
</tr>
<tr>
<td>Setting</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hospital</td>
<td>7</td>
<td>6</td>
<td>2</td>
<td>15</td>
</tr>
<tr>
<td>Community</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Not known</td>
<td>12</td>
<td>6</td>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td>Affected joint</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knee</td>
<td>14</td>
<td>13</td>
<td>1</td>
<td>28</td>
</tr>
<tr>
<td>Hip</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Hand</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Multiple/not specified</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>QST modality†</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mechanical</td>
<td>18</td>
<td>8</td>
<td>2</td>
<td>28</td>
</tr>
<tr>
<td>Pressure</td>
<td>13</td>
<td>6</td>
<td>2</td>
<td>21</td>
</tr>
<tr>
<td>Puncture, brush</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Vibratory</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Electrical</td>
<td>4</td>
<td>6</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>Thermal</td>
<td>7</td>
<td>4</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>Chemical</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Outcome measure†</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pain threshold</td>
<td>17</td>
<td>12</td>
<td>1</td>
<td>30</td>
</tr>
<tr>
<td>Detection threshold</td>
<td>9</td>
<td>1</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>Pain intensity</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>Muscle reflex</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Brain activity</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Other†</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>QST test sites‡</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affected joint</td>
<td>11</td>
<td>8</td>
<td>2</td>
<td>21</td>
</tr>
<tr>
<td>Distal to affected joint</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Remote from affected joint</td>
<td>9</td>
<td>4</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>Other‡</td>
<td>6</td>
<td>5</td>
<td>1</td>
<td>12</td>
</tr>
</tbody>
</table>

Affected joint – main location of OA.
† Randomised control trial, uncontrolled trial.
‡ One study may involve more than one QST modality, outcome measure and test site.
§ Pain area, duration, tolerance.
± Average threshold reading from a number of anatomical test sites, or the affected joint or anatomical test site not specified and thus not categorised.

The next most common modalities were electrical (11/41) and thermal (11/41) and a small number of studies (2/41) applied chemical stimulus.

The studies were primarily concerned with pain thresholds (30/41) and other common outcome measures included detection, i.e., first sensation thresholds (10/41) and intensity of pain measured on a scale (9/41). Small numbers of studies (four and three, respectively, of 41) examined muscle reflexes using electromyographic (EMG) recording, and brain activity during QST using functional Magnetic Resonance Imaging (fMRI) or Positron Emission Tomography (PET) scanning.

The studies typically involved QST of more than one anatomical site, and the largest number of sites included in a single study was 24.22 Coding of the anatomical sites demonstrated that 21 of the 41 studies applied stimuli to the primary site affected by OA, and that 13 studies tested remote and eight studies distal sites. Twelve studies measured the average threshold from two or more sites including the affected joint. Overall, differences in study protocols resulted in heterogeneity which limited the comparison of study results even within a single modality.

Reliability of QST

Three papers20,21,37 reported ICCs for the test–retest reliability of QST in OA. In a knee OA study by Moss et al.,37 the time between PPT measurements was at least 48 h and the ICCs were near perfect with an ICC of 0.98 (95%CI 0.96, 0.99) for both the affected knee and the distal ipsilateral heel. Wessel20 tested PPTs in knee OA on six sites on or above the knee in both legs. Measurements were taken 5–10 days apart and the ICCs varied from adequate (0.58) to excellent (0.91). The most comprehensive analysis of ICCs was carried out in a knee OA study by Wylde et al.,31 which involved two types of mechanical (light touch, pressure pain) and three types of thermal stimuli (cold, warm, hot pain). PPTs were found to be the least variable measurement over a 1-week period, and the ICCs were excellent for the affected knee (0.83, 95%CI 0.72, 0.90) and the contralateral knee (0.77, 95%CI 0.63, 0.86) and the forearm (0.86, 95%CI 0.77, 0.92).

Ability of QST to differentiate between groups

Of 41 studies included, 13 reported that people with OA were more sensitive than normal controls to painful stimuli.
(PPTs\textsuperscript{7,20–26,38}, mechanical and thermal pain thresholds\textsuperscript{45}, thermal pain threshold\textsuperscript{46}, punctate pain threshold\textsuperscript{29,47}, or chemical pain rating\textsuperscript{27}). Two further studies\textsuperscript{28,48} applying noxious electrocutaneous stimuli reported that the threshold to elicit flexor withdrawal reflex was significantly lower in the OA group. Three studies reported no significant difference between the OA and the normal control groups (‘piston’ pressure pain ratings\textsuperscript{31}, finger pressure pain ratings\textsuperscript{49}, heat and cold pain\textsuperscript{47}). In contrast to pain studies, sensory detection thresholds in people with OA were reported to be either higher, i.e., indicative of hypoesthesia (vibratory\textsuperscript{50,51}, innocuous punctate\textsuperscript{29}), or similar to those of healthy controls (electrical\textsuperscript{46}, light touch and innocuous cold\textsuperscript{47}).

Sufficient data were available from seven studies to undertake meta-analysis comparing PPTs between people with OA and healthy controls. The mean (standard deviation (SD)) of the PPT in kg/cm\textsuperscript{2} for different anatomical test sites among the included studies varied from 1.81 (1.00) to 5.22 (2.26) in people with OA, and from 3.40 (0.84) to 11.20 (2.03) in healthy controls. SMDs between the groups ranged from –0.47 (95%CI –1.00, 0.06) to –3.04 (95%CI –3.77, –2.31) (Fig. 2), where the negative value means lower pain threshold in OA. All but one site in one study\textsuperscript{22} demonstrated a significant lower PPT in OA. The one study\textsuperscript{22} that reported that PPTs were higher in people with OA than healthy controls was not included in the meta-analysis because the affected joint was not specified and therefore the anatomical test sites could not be categorised in relation to the affected joint.

The pooled SMD, calculated by selecting the anatomical QST site with the smallest SMD from each study, was –0.87 (95%CI –1.08, –0.66) (Fig. 3). Funnel plot and Egger’s test (bias = –0.70, P = 0.69) from the seven studies did not suggest significant publication bias (Fig. 4), and the Q test (Q = 6.34, P = 0.39) and the I\textsuperscript{2} test (5%, 95%CI 0%, 61%) did not indicate heterogeneity between studies (Fig. 3).

**QST results from different anatomical test sites**

For most studies, PPTs were reported to be reduced in people with OA compared with controls, irrespective of whether tested at affected, distal or remote sites. This is exemplified by Imamura et al.\textsuperscript{22} who reported significantly lower PPTs in people with knee OA when applying stimuli to skin overlying muscle or bone at all 18 anatomical test sites including upper thigh and lower back. Similarly, Lee et al.\textsuperscript{23} reported lower thresholds for thumb, shoulder and front of the thigh, and Wylde et al.\textsuperscript{38} for forearm, in people with knee OA. However, one study\textsuperscript{24} reported that PPTs were higher in people with OA than normal healthy controls and another\textsuperscript{24} reported that PPTs for people with hip OA on sites contralateral to the affected site were similar to those of normal healthy controls.

The smallest SMD for the affected joint, distal and the remote anatomical test sites were selected from each study and pooled within these three categories (Fig. 5). The SMD was larger for the affected joint sites (SMD = –1.24, 95%CI –1.54, –0.93) compared with remote sites (SMD = –0.88, 95%CI –1.11, –0.65). The
minimum mean (SD) of the PPTs (kg/cm²) in the OA group was 2.01 (1.22) compared to 3.40 (0.84) in the control group for the affected sites, whereas it was 2.27 (1.52) vs 3.45 (1.00) for the remote sites. For distal sites, which involved only three anatomical sites from two studies, there was no significant difference between the OA and the control group.

Other analyses

The minimum sample size needed to assure adequate statistical power to detect differences between the two groups was investigated for PPTs. For the affected joint the minimum SMD was −0.68 (mean difference −1.78 kg/cm², SD 2.57). The sample size needed to

---

### AFFECTED JOINT

- **Arendt-Nielsen 2010**
  - Lateral side of the knee: -0.84 (-1.39, -0.29)
  - Patellar tendon: -0.80 (-1.36, -0.25)
  - Medial side of the knee: -0.68 (-1.22, -0.14)

- **Imamura 2008**
  - Patellar tendon: -3.04 (-3.77, -2.31)
  - Rectus femoris muscle: -2.54 (-3.21, -1.87)
  - Vastus medialis muscle: -2.53 (-3.20, -1.86)
  - Pes anserinus bursae: -2.11 (-2.73, -1.48)
  - Vastus lateralis muscle: -1.73 (-2.32, -1.14)
  - Popliteus muscle: -1.24 (-1.79, -0.68)

- **Kosek 2000b**
  - Site of max on-going pain: -1.56 (-2.41, -0.71)

- **Wessel**
  - Lateral joint line: -2.28 (-3.12, -1.44)
  - Medial ligament: -2.11 (-2.93, -1.29)
  - Vastus medialis: -1.79 (-2.56, -1.01)
  - Medial joint line: -1.72 (-2.48, -0.96)
  - Lateral ligament: -1.66 (-2.42, -0.91)
  - Vastus lateralis: -1.60 (-2.35, -0.85)

- **Wyde 2011b**
  - Index knee: -1.34 (-1.71, -0.98)

### DISTAL

- **Arendt-Nielsen 2010**
  - Tibialis anterior: -0.47 (-1.00, 0.06)

- **Imamura 2008**
  - Peroneus longus muscle: -2.22 (-2.86, -1.59)
  - Tibialis anterior muscle: -1.98 (-2.59, -1.36)

### REMOTE

- **Arendt-Nielsen 2010**
  - Arm: -0.83 (-1.37, -0.28)

- **Imamura 2008**
  - L4: -2.36 (-3.01, -1.71)
  - L3: -2.34 (-2.99, -1.69)
  - L5: -1.99 (-2.60, -1.37)
  - S2: -1.99 (-2.60, -1.37)
  - L2: -1.92 (-2.52, -1.31)
  - Adductor longus: -1.65 (-2.23, -1.07)
  - L1: -1.63 (-2.22, -1.05)
  - S1: -1.56 (-2.14, -0.99)
  - Quadratus lumborum muscle: -1.30 (-1.85, -0.74)
  - Iliacus muscle: -1.00 (-1.54, -0.46)

- **O’Driscoll 1975**
  - Contralateral site: -0.79 (-1.56, -0.02)
  - Forehead: -0.74 (-1.36, -0.11)

- **Wyde 2011**
  - Forearm: -0.91 (-1.26, -0.56)

### OTHER

- **Lee 2011**
  - First MCP joint: -1.05 (-1.59, -0.50)
  - Trapezius muscle: -0.84 (-1.38, -0.30)
  - Quadriceps muscle: -0.80 (-1.33, -0.26)

---

**Fig. 2.** Forest plot showing the SMD (95% CIs) for PPT between people with OA and healthy controls, grouped by QST site in relation to the affected joint (affected joint, distal, remote, other). L1−L5 = lumbar vertebrae, S1−S2 = sacral vertebrae, MCP = metacarpophalangeal. Other = average threshold measured from both sides of the body. Negative SMD means lower pain threshold in OA compared with control.
detect this difference for a future QST study on this anatomical site would be 45 people per group to give a power of 90% and false positive error of less than 5% (Fig. 6). The minimum SMD for sites distal to the affected joint was −0.47 (mean difference −0.98 kg/cm², SD 2.03), and the sample size needed to detect this difference would be 91 per group. For sites remote from the affected joint the minimum SMD was −0.74 (mean difference −0.73 kg/cm², SD 0.97)²⁵,²⁶, and the sample size required to detect this difference would be 39 per group. Finally, for a study which aims to test all three sites (affected, distal and remote) the minimum sample size would be 91 people per group.

**Discussion**

**Application and reliability of QST as a research method**

Pressure stimulus emerged as the most common method for assessing somatosensory responses and for showing abnormalities in the target population, and QST was more often applied to the affected joint than distal and remote sites. The magnitude of difference (SMD) in PPTs was large, and the minimum number required to differentiate people with OA and healthy controls using QST at a single site was 45 per group to ensure 90% power and less than 5% false positive error. High test—retest repeatability indicates that PPTs behave as stable characteristics over periods at least up to 1 week, and that pressure stimulus provides a reliable tool for measuring pain thresholds.

![Forest plot showing the SMD (95%CI) for PPT between people with OA and healthy controls. The anatomical test site with the smallest SMD was selected from each study. Test for heterogeneity: Cochran Q = 6.34, P = 0.39, I² (inconsistency) = 5% (95% CI 0—61%). Negative SMD means lower pain threshold in OA compared with control.](image)

**Ability of QST to differentiate between groups**

The findings from this review indicate that pain threshold may be lower in people with OA than healthy controls. However, the results from the included studies are variable and sometimes contradictory. While some studies showed that people with OA may have lower pain²⁷,²⁸,³¹,³⁴,³⁶,³⁸ and higher sensory detection thresholds, others found that pain²⁹,³⁰,³¹,³³ and sensory detection thresholds in people with OA were similar to those of healthy controls. PPTs measured over or remote from the affected joint displayed similar abilities to differentiate between groups.

**Anatomical sites of QST and evidence of sensitisation**

The key finding from the meta-analysis was that compared with healthy controls, people with OA had lower PPTs not only in the affected joint but also at remote sites not directly affected by OA (spreading sensitisation). Investigating the locations and patterns of pain — such as categorising knee OA pain as localised, regional or diffuse²⁴ — holds promise for better understanding sensory abnormalities. Low PPTs over the affected joint may represent nociception, for example due to local inflammation, while reduced pain thresholds at sites remote from the affected joint support the hypothesis that central sensitisation or reduced descending inhibitory control may be important mediators of chronic musculoskeletal pain.¹⁶ Pain experienced distal to the site of pathology is described as ‘referred pain’, and increased brain activation during stimulation of sites corresponding to referred pain areas has been associated with neuropathic-like pain symptoms, and is further
suggestive of central sensitisation. Associations between altered pain thresholds and brain function, as determined by fMRI, and normalisation after surgical interventions that reduce nociceptive input further support the hypothesis that QST results reflect alterations in central pain processing.

Central sensitisation both may be a consequence of ongoing nociceptive input and a mechanism by which OA pain is maintained. Normalisation of PPTs following knee replacement surgery indicates that treatments directed at the joint may reduce sensitisation, whereas poor surgical outcomes in people with QST evidence of preoperative central sensitisation suggests persistent pain processing abnormalities. Further prospective studies would be required to determine whether evidence of somatosensory abnormalities could be used clinically to predict persistent post-operative pain, or to help target adjunctive therapies in order to improve pain outcomes after surgery, as recently was suggested for the centrally acting analgesic pregabalin.

Methodological issues

Bearing in mind the diversity of QST protocols we observed across studies, in future research it may be necessary to improve the reproducibility and comparability of studies. The adoption of standardised QST protocols (e.g., the German Research Network on Neuropathic Pain) in OA studies would facilitate comparisons between studies and groups, and pooling of data would permit greater confidence in the generalisability of study results. Generalisability can also be improved by appropriate sample size and power calculation.

Further, methodological quality could be improved by experimental designs that link up with theories of somatosensory processing. Identification of the locations and patterns of pain, and exploring the relationship between pain patterns and OA risk factors, will provide valuable information about the aetiology and nature of sensory abnormalities. Using QST alongside other clinical study methods (e.g., imaging evidence) may be particularly useful in studying the complex mechanisms of peripheral and central sensitisation. Methods have been developed to enable comparison of individual patients with OA to the group average of healthy participants, thus complementing the current practice of comparing somatosensory abnormalities at group level. Finally, large RCTs are required to assess the ability of QST to phenotype individual patients in order to direct therapy and improve treatment outcomes.

Caveats of the study

A number of caveats need to be noted in this systematic review. Firstly, the screening and selection of studies were carried out by only one assessor, which means that some relevant studies may have been excluded from the review. Secondly, quality assessment was only carried out for the studies included in the meta-analysis, and the brief narrative review of findings was not interpreted in context with the methodological quality of the studies. Thirdly, it is possible that OA was present in joints that were not examined and reported in the studies, which would affect the precision of our analysis. Fourthly, the small number of studies for each QST modality and the heterogeneity of study methods (e.g., imaging evidence) may be particularly useful in studying the complex mechanisms of peripheral and central sensitisation. Methods have been developed to enable comparison of individual patients with OA to the group average of healthy participants, thus complementing the current practice of comparing somatosensory abnormalities at group level. Finally, large RCTs are required to assess the ability of QST to phenotype individual patients in order to direct therapy and improve treatment outcomes.

Conclusion

Current evidence confirms that people with OA have lower PPTs. This can be detected at both affected and unaffected sites, suggesting that central sensitisation contributes to pain in OA. QST merits further investigation as a research tool to help understand pain mechanism in OA. More research is also needed to realise the potential of QST to define pain phenotypes that could help target specific treatments to those who are most likely to benefit.

Author contributions

Project conceptualisation WZ and DAW.
Data study design: WZ, AS, DAW, DMcW.
Data collection/validation: AS, DAW, DMcW, LC, BM and WZ.
Data analysis: AS, DAW, DMcW and WZ.
Result interpretation: AS, DAW, DMcW, LC, BM, VW, LAN and WZ.
Reporting & editing: AS, DAW, DMcW, LC, BM, VW, LAN and WZ.
Final approval of the version to be submitted: AS, DAW, DMcW, LC, BM, VW, LAN and WZ.
Project guarantor: WZ.

Funding

There was no external funding received in support of this systematic review.

Conflicts of interest

The authors declare that they have no competing interests.

Acknowledgements

We thank Dr Henryk Faas and Professor Leonardo Punzi for their help in reviewing foreign language articles. We are also grateful to Professor Dorothee Auer, Professor Victoria Chapman, Professor Michael Doherty, Dr Sara Kelly, Professor Nadina Lincoln, and Professor Brigitte Scammell for their guidance and encouragement during this project.

Appendix 1. MEDLINE search strategy

1. osteoarthritis.mp. or exp Osteoarthritis, Hip/ or exp Osteoarthritis/ or exp Osteoarthritis, Spine/ or exp Osteoarthritis, Knee/
2. osteoarthrosis.mp.
3. gonarthrosis.mp.
4. gonitis.mp.
5. coxarthrosis.mp.
6. coxitis.mp.
7. knee pain.mp.
8. exp Osteophyte/ or osteophyte*.mp.
9. (joint space adj6 narrow*).tw.
10. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11
12. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11
13. quantitative sensory.mp.
14. exp Sensory Thresholds/
15. exp Pain Threshold/
16. threshold*.mp.
17. 13 or 14 or 15 or 16
18. 12 and 17
Appendix 2. Checklist for study quality

### Scoring: yes = 1 no = 0; unless stated otherwise

#### Reporting
1. Does the study provide a clear hypothesis/aim/objective for measurement of pain/sensory/reflex thresholds using QST* to detect differences in thresholds — why, for what purpose.
2. Are the main outcomes to be measured clearly described in the Introduction or Methods section (and not just in the Results): thresholds and ratings.
3. Are the characteristics of the patients included in the study clearly described:
   - cohorts and trials:
     - inclusion criteria yes = 0.5;
     - exclusion criteria yes = 0.5.
   - case control studies:
     - case-definition yes = 0.5;
     - source for controls yes = 0.5.
   - case only studies: case-definition yes = 1.
4. Are the distributions of principal confounders in each group of subjects to be compared clearly described: patient demographics for age and gender either in a table or in text.
5. Are the main findings of the study clearly described by providing simple outcome data: numeric values (e.g., mean value) so that reader can check the major analyses and conclusions.
6. Have actual probability values been reported for the main outcomes (e.g., 0.035 rather than <0.05) except where P < 0.001.

#### Validity
8. If any of the results were based on ‘data dredging’, was this made clear (if no data dredging reported then answer yes).
9. Were the statistical tests used to assess the main outcomes appropriate: to detect differences in thresholds between time-points, groups, interventions etc.
10. Were the cases representative of the entire population from which they were recruited: unselected sample of consecutive patients or a random sample.
11. Were power calculations carried out for the primary outcome: threshold measurement.
12. Was there adequate adjustment for confounding in the analyses from which the main findings were drawn: were the groups age and gender matched — if not, was this caveat discussed.

**Total**

* QST defined as inclusion of one or several of the following modalities: mechanical (e.g., pressure, punctate, and light touch), thermal, electrical, chemical.

### Appendix 3. Individual study characteristics

<table>
<thead>
<tr>
<th>Author &amp; year</th>
<th>Country</th>
<th>Design</th>
<th>Setting</th>
<th>Participants</th>
<th>Age mean (SD/range)</th>
<th>Female%</th>
<th>Affected joint</th>
<th>QST modalities</th>
<th>QST affected</th>
<th>Distal</th>
<th>Remote</th>
<th>Other*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arendt-Nielsen et al. (2010)</td>
<td>Denmark</td>
<td>cc</td>
<td>Unknown</td>
<td>72</td>
<td>62 (8)</td>
<td>50</td>
<td>Knee</td>
<td>C, M</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Bajaj et al. (2001)</td>
<td>Denmark</td>
<td>cc</td>
<td>Unknown</td>
<td>28</td>
<td>40 (12)</td>
<td>57</td>
<td>Knee and below</td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brucini et al. (1981)</td>
<td>Italy</td>
<td>cc</td>
<td>Unknown</td>
<td>26</td>
<td>(24–76)</td>
<td>50</td>
<td>Knee</td>
<td>E</td>
<td></td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Buffington et al. (2005)</td>
<td>USA</td>
<td>cc</td>
<td>Comm</td>
<td>16</td>
<td>48 (13/32–76)</td>
<td>44</td>
<td>Knee</td>
<td>M</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Courtney et al. (2009)</td>
<td>USA</td>
<td>cc</td>
<td>Comm</td>
<td>40</td>
<td>61 (9/45–75)</td>
<td>60</td>
<td>Knee</td>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Courtney et al. (2010)</td>
<td>USA</td>
<td>cc</td>
<td>Hosp</td>
<td>20</td>
<td>60 (45–75)</td>
<td>60</td>
<td>Knee</td>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creamer et al. (1998)</td>
<td>USA</td>
<td>co</td>
<td>Hosp</td>
<td>58</td>
<td>66 (10)</td>
<td>69</td>
<td>Knee</td>
<td>E</td>
<td></td>
<td>M</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>Creamer et al. (1999)</td>
<td>USA</td>
<td>co</td>
<td>Hosp</td>
<td>68</td>
<td>66 (10/40–86)</td>
<td>69</td>
<td>Knee</td>
<td>M</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creamer et al. (1999)</td>
<td>USA</td>
<td>co</td>
<td>Unknown</td>
<td>37</td>
<td>65 (9)</td>
<td>59</td>
<td>Knee</td>
<td>M</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Defrin et al. (2005)</td>
<td>Israel</td>
<td>co</td>
<td>Hosp</td>
<td>62</td>
<td>69 (9/55–75)</td>
<td>50</td>
<td>Knee</td>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dragani et al. (1992)</td>
<td>Italy</td>
<td>cc</td>
<td>Unknown</td>
<td>20</td>
<td>62 (55–72)</td>
<td>100</td>
<td>Knee</td>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emery et al. (2006)</td>
<td>USA</td>
<td>co</td>
<td>Unknown</td>
<td>62</td>
<td>63 (8/50–76)</td>
<td>56</td>
<td>Knee</td>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Farrell et al. (2000)</td>
<td>Australia</td>
<td>cc</td>
<td>Unknown</td>
<td>80</td>
<td>65 (8)</td>
<td>79</td>
<td>Hand</td>
<td>M, T</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Farrell et al. (2000)</td>
<td>Australia</td>
<td>co</td>
<td>Unknown</td>
<td>24</td>
<td>64 (7)</td>
<td>100</td>
<td>Hand</td>
<td>E, M, T</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>France et al. (2004)</td>
<td>USA</td>
<td>co</td>
<td>Comm</td>
<td>132</td>
<td>63 (8)</td>
<td>56</td>
<td>Knee</td>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gerecz-Simon et al. (1989)</td>
<td>Canada</td>
<td>cc</td>
<td>Unknown</td>
<td>72</td>
<td>48 (20–70)</td>
<td>50</td>
<td>Hip/knee</td>
<td>M</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gwilym et al. (2009)</td>
<td>UK</td>
<td>cc</td>
<td>Hosp</td>
<td>24</td>
<td>64 (9)</td>
<td>46</td>
<td>Hip</td>
<td>M, T</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hendiani et al. (2003)</td>
<td>USA</td>
<td>cc</td>
<td>Hosp</td>
<td>55</td>
<td>49 (10/20–73)</td>
<td>84</td>
<td>Knee</td>
<td>M</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Imamura et al. (2008)</td>
<td>Brazil</td>
<td>cc</td>
<td>Unknown</td>
<td>84</td>
<td>71 (7)</td>
<td>100</td>
<td>Knee</td>
<td>M</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Keele et al. (1997)</td>
<td>USA</td>
<td>co</td>
<td>Hosp</td>
<td>40</td>
<td>62 (19)</td>
<td>53</td>
<td>Knee</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kosek and Ordeberg (2000)</td>
<td>Sweden</td>
<td>cc</td>
<td>Unknown</td>
<td>28</td>
<td>52 (29–67)</td>
<td>64</td>
<td>Hip</td>
<td>M, T</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Krause et al. (1995)</td>
<td>Germany</td>
<td>cc</td>
<td>Unknown</td>
<td>100</td>
<td>53</td>
<td>58</td>
<td>Knee</td>
<td>E, T</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix 3. (continued)

<table>
<thead>
<tr>
<th>Author &amp; year</th>
<th>Country</th>
<th>Design</th>
<th>Setting</th>
<th>Participants</th>
<th>Age mean (SD/range)</th>
<th>Female%</th>
<th>Affected joint</th>
<th>QST modalities</th>
<th>QST affected</th>
<th>Distal</th>
<th>Remote</th>
<th>Other*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kulkarni et al. (2007)³⁰</td>
<td>UK</td>
<td>co</td>
<td>Unknown</td>
<td>12</td>
<td>59 (52–67)</td>
<td>50</td>
<td>Knee</td>
<td>T</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lee et al. (2011)³³</td>
<td>USA</td>
<td>cc</td>
<td>Comm</td>
<td>59</td>
<td>58 (9)</td>
<td>73</td>
<td>Knee</td>
<td>M, T</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lundblad et al. (2008)³⁶</td>
<td>Sweden</td>
<td>co</td>
<td>Hosp</td>
<td>69</td>
<td>68 (40–80)</td>
<td>51</td>
<td>Knee</td>
<td>E</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Martinez et al. (2007)³⁴</td>
<td>USA</td>
<td>co</td>
<td>Hosp</td>
<td>20</td>
<td>69 (9)</td>
<td>95</td>
<td>Knee</td>
<td>M, T Y</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moss et al. (2007)³¹</td>
<td>Australia</td>
<td>cc</td>
<td>Comm</td>
<td>38</td>
<td>63 (11/40–87)</td>
<td>66</td>
<td>Knee</td>
<td>M</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O’Driscoll and Jayson</td>
<td>(1974, 1975)²⁵,²⁶</td>
<td>UK</td>
<td>cc</td>
<td>Unknown</td>
<td>55</td>
<td>68</td>
<td>Hip</td>
<td>M</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oliveira et al. (2009)³⁰</td>
<td>Portugal</td>
<td>cc</td>
<td>Unknown</td>
<td>86</td>
<td>55 (10)</td>
<td>78</td>
<td>Not specified</td>
<td>M</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sayed-Noor et al. (2008)³¹</td>
<td>Sweden</td>
<td>cc</td>
<td>Hosp</td>
<td>36</td>
<td>65 (45–82)</td>
<td>83</td>
<td>Hip</td>
<td>M</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Schmidt et al. (2010)²²</td>
<td>Germany</td>
<td>tri</td>
<td>Hosp</td>
<td>30</td>
<td>(44–68)</td>
<td>73</td>
<td>Hand/hip/knee</td>
<td>M</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shakoor et al. (2008)³⁰</td>
<td>USA</td>
<td>cc</td>
<td>Hosp</td>
<td>41</td>
<td>52 (13)</td>
<td>76</td>
<td>Knee</td>
<td>M</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Takeda and Wessel (1994)³⁵</td>
<td>USA</td>
<td>cc</td>
<td>Comm</td>
<td>27</td>
<td>62 (10)</td>
<td>41</td>
<td>Hip</td>
<td>M</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Takeda and Wessel (1995)²⁰</td>
<td>Canada</td>
<td>RCT</td>
<td>Unknown</td>
<td>40</td>
<td>62 (9)</td>
<td>50</td>
<td>Knee</td>
<td>M</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wilder-Smith et al. (2001)³⁰</td>
<td>South Africa</td>
<td>RCT</td>
<td>Hosp</td>
<td>87</td>
<td>59 (55–65)</td>
<td>32</td>
<td>Hip/knee</td>
<td>E</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Williams et al. (2004)³⁷</td>
<td>USA</td>
<td>cc</td>
<td>Comm</td>
<td>154</td>
<td>66 (9)</td>
<td>66</td>
<td>Knee</td>
<td>M</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wyld et al. (2011)³³</td>
<td>UK</td>
<td>cc</td>
<td>Hosp</td>
<td>100</td>
<td>70 (8)</td>
<td>44</td>
<td>Knee</td>
<td>M, T Y</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wyld et al. (2011)³³</td>
<td>UK</td>
<td>cc</td>
<td>Hosp</td>
<td>100</td>
<td>70 (8)</td>
<td>44</td>
<td>Knee</td>
<td>M, T Y</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Study design: cc = case control, co = cases only, tri = uncontrolled trial.
Setting: Comm = community-based, Hosp = hospital-based.
QST modalities: C = Chemical, E = Electrical, M = mechanical, T = Thermal.
* Other for QST test sites includes averaged (e.g., both sides of the body) and non-specified sites.

References


