Pain Sensitivity Risk Factors for Chronic TMD: Descriptive Data and Empirically Identified Domains from the OPPERA Case Control Study

Joel D. Greenspan[1], **Gary D. Slade**[2,3,4], **Eric Bair**[4,5,6], **Ronald Dubner**[1], **Roger B. Fillingim**[7], **Richard Ohrbach**[8], **Charlie Knott**[9], **Flora Mulkey**[5], **Rebecca Rothwell**[4], and **William Maixner**[4,6]

[1]Department of Neural and Pain Sciences, and Brotman Facial Pain Center, University of Maryland Dental School, Baltimore, MD

[2]Department of Dental Ecology, University of North Carolina at Chapel Hill, Chapel Hill, NC

[3]Department of Epidemiology, University of North Carolina at Chapel Hill Chapel Hill, NC

[4]Center for Neurosensory Disorders, University of North Carolina at Chapel Hill Chapel Hill, NC

[5]Department of Biostatistics, University of North Carolina at Chapel Hill Chapel Hill, NC

[6]Department of Endodontics, University of North Carolina at Chapel Hill Chapel Hill, NC

[7]Department of Community Dentistry & Behavioral Science, University of Florida, Gainesville, FL

[8]Department of Oral Diagnostic Services, University at Buffalo, Buffalo, NY

[9]Battelle Memorial Institute, Durham, NC

Abstract

Many studies report that people with temporomandibular disorders (TMD) are more sensitive to experimental pain stimuli than TMD-free controls. Such differences in sensitivity are observed in remote body sites as well as in the orofacial region, suggesting a generalized upregulation of nociceptive processing in TMD cases. This large case-control study of 185 adults with TMD and 1,633 TMD-free controls measured sensitivity to painful pressure, mechanical cutaneous, and heat stimuli, using multiple testing protocols. Based on an unprecedented 36 experimental pain measures, 28 showed statistically significantly greater pain sensitivity in TMD cases than controls. The largest effects were seen for pressure pain thresholds at multiple body sites and cutaneous mechanical pain threshold. The other mechanical cutaneous pain measures and many of the heat pain measures showed significant differences, but with lesser effect sizes. Principal component analysis (PCA) of the pain measures derived from 1,633 controls identified five components labeled: (1) heat pain ratings, (2) heat pain aftersensations and tolerance, (3) mechanical cutaneous

Correspondence: Joel D. Greenspan, Dept. of Neural and Pain Sciences, 8 South, Univ. of Maryland Dental School, Baltimore MD 21201, 410-706-7090 (Telephone), 410-706-0865 (FAX), jgreenspan@umaryland.edu.

pain sensitivity, (4) pressure pain thresholds, and (5) heat pain temporal summation. These results demonstrate that, compared to TMD-free controls, chronic TMD cases are more sensitive to many experimental noxious stimuli at extra-cranial body sites, and provides for the first time the ability to directly compare the case-control effect sizes of a wide range of pain sensitivity measures.

Keywords

QST; chronic facial pain; heat pain; pressure pain; mechanical pain

INTRODUCTION

A number of studies have reported chronic TMD cases to be more sensitive than pain-free controls to experimental pain stimuli, using a variety of protocols and pain measures.[23–25,32,37] However, such differences are not observed in all studies.[3,5,37] Among studies showing significant case-control differences, chronic TMD cases often are more pain sensitive in remote body sites as well as in the orofacial region, suggesting a generalized upregulation of nociceptive input processing in this patient population.

It is not clear which circumstances result in finding significant differences between TMD cases and controls in experimental pain sensitivity. In previous studies, various modes of assessing experimental pain sensitivity revealed differences between TMD cases and controls, including mechanical and heat stimuli, as well as cutaneous and deep tissue stimuli. Case-control differences also have been observed using assessments of threshold, tolerance, suprathreshold ratings, and temporal summation (TS) of pain.[33] However, there is currently no indication that a particular type of stimulus or assessment method is more or less likely to identify case-control differences in experimental pain sensitivity.

Yet, important differences exist among various types of stimuli and among different assessment protocols. Different modalities of experimental pain stimulation (i.e., mechanical, thermal) recruit different groups of nociceptors and activate only partially overlapping CNS pathways. Furthermore, different pain assessment measures (i.e., thresholds, suprathreshold ratings, TS) engage different aspects of pain perception and thus, at some level, different CNS processes.[1] Hence, there is only modest correlation between the same types of assessments of different stimulus modalities, or between different types of assessments of the same stimulus modality.[2,17,18,21] This suggests that conditions affecting experimental pain sensitivity may have differential effects upon components of the pain processing system, showing varying results depending upon the experimental protocol and the particular aspect of pain evaluated.

The Orofacial Pain: Prospective Evaluation and Risk Assessment (OPPERA) study is a prospective cohort study designed to identify causal determinants of TMD pain. This study is based on a conceptual model in which enhanced pain sensitivity represents a critically important intermediate phenotypic risk factor for development of TMD.[8] The findings presented below describe the baseline assessment of pain sensitivity in a large group of TMD-free controls as well as a cohort of TMD cases. A wide range of quantitative sensory testing (QST) was conducted in order to assess multiple aspects of experimental pain sensitivity. While previous research has found that some aspects of pain sensitivity are associated with TMD, no studies have assessed this characteristic with the broad array of measures used here. Furthermore, in addition to case-control comparisons of individual QST measures, we have conducted principal component analysis to characterize latent constructs of pain sensitivity.

METHODS

Study setting and participants

As described elsewhere,[34] the OPPERA baseline case-control study used advertisements, emails, flyers and word-of-mouth to recruit people who had chronic TMD ("cases") and people who did not ("controls"). They were recruited between May 2006 and November 2008 from communities in and around academic health centers at four US study sites: Baltimore, MD; Buffalo, NY; Chapel Hill, NC; and Gainesville, FL. At each study site, the target was to recruit 800 controls and variable numbers of cases based on local operational requirements, for a total of 3,200 controls and 200 cases. The actual number enrolled was 3,263 controls and 185 cases.

The classification of TMD was based on the Research Diagnostic Criteria for Temporomandibular Disorder.[9] In summary, cases met all three of the following criteria: during the telephone interview, (i) pain reported with sufficient frequency in the cheeks, jaw muscles, temples or jaw joints during the preceding six months (at least 15 days in the preceding month and at least five days per month in each of the five months preceding that); during the examination, (ii) pain reported in the examiner-defined orofacial region for at least 5 days out of the prior 30 days; and (iii) pain reported in at least three masticatory muscles or at least one temporomandibular joint in response to palpation of the orofacial muscles or maneuver of the jaw. Examiners defined the orofacial region by touching the following anatomical areas bilaterally: temporalis, preauricular, masseter, posterior mandibular, and submandibular. Controls met all six of the following criteria: during the telephone interview, (i) pain reported infrequently in the cheeks, jaw muscles, temples or jaw joints (no orofacial pain in the preceding month and no more than four days per month in any of the five months preceding that); (ii) no more than four headaches per month within the preceding three months; (iii) never diagnosed with TMD; (iv) no use of night guard occlusal splint; and during the examination, (v) pain reported in the examiner-defined orofacial region for no more than 4 days in the prior 30 days; (vi) classified as having neither myalgia nor arthralgia. However, controls could be positive or negative with respect to pain in response to palpation or jaw maneuver. Additional study-wide criteria for all study participants were: aged 18–44 years; fluent in English; negative responses to each of 10 questions regarding significant medical conditions; no history of facial injury or surgery; not receiving orthodontic treatment; not pregnant or nursing.

This analysis uses data from all 185 recruited TMD cases and one half of the 3,263 recruited controls (1,633 people). The controls for this analysis were selected at random so that data from people in the reserved sample could be used for validation studies that will be reported elsewhere. The accompanying paper gives a more detailed account of study recruitment, case-classification methods and inclusion and exclusion criteria.[34]

The OPPERA study was reviewed and approved by institutional review boards at each of the four study sites and at the data coordinating center, Battelle Memorial Institute. All study participants verbally agreed to a screening interview done by telephone, and they provided informed, signed consent for all other study procedures.

Psychophysical protocols

Quantitative sensory testing was conducted in three sensory domains, in the following order: pressure pain, mechanical cutaneous (pricking) pain, and heat pain.

Pressure pain. Pressure pain thresholds (PPT) were assessed using a commercially available pressure algometer (Somedic; Hörby, Sweden). Five body sites were tested, bilaterally, in the following order: (1) the center of the temporalis muscle, (2) the center of the masseter

muscle, (3) overlying the TMJ, (4) the center of the trapezius muscle, and (5) overlying the lateral epicondyle. The protocol involved manual application of the algometer, with which the examiner would increase pressure at a steady rate (30 kPa/s), until the participant indicated first pain sensation by pressing a button. The first trial at each test site was considered a practice trial, and was excluded from data analysis. In subsequent trials, the pressure level at the time of the button press was recorded as a threshold estimate. If no response was given at the point the stimulus reached 600kPa, a value of 600 was used as the threshold value. This procedure was repeated at the same test site until either (1) two values were recorded within 20kPa of one another, or (2) five trials were administered. In either case, the mean of the two closest values was recorded as the threshold estimate.

The examiners from each of the four study sites participated in a reliability exercise at an early stage of this project. PPTs were assessed on a group of 16 subjects (not part of the main study), each of whom was tested by two examiners at each of five body sites. The overall ICC was 0.91, ranging from 0.87 – 0.94.

Mechanical cutaneous (pinprick) pain—Pricking pain sensitivity was assessed using a set of weighted probes, manufactured locally, matching those used by the German Neuropathic Pain Network.[31] This set of probes had a flat contact area of 0.2 mm diameter, and exerted forces between 8 and 512 mN. Stimuli were applied to the dorsum of the digits 2–4. Measures included pain threshold, ratings of pain intensity in response to the two largest stimulus intensities, and TS of pain. Pain threshold was derived using an adaptive staircase method,[6] calculated as the geometric mean of five series of ascending and descending stimulus intensities. If subjects gave two "No" responses in a row using the 512mN probe, the staircase was halted and a value of 512 was used as the threshold value.

After threshold determination, suprathreshold cutaneous mechanical pain sensitivity was assessed using a protocol similar to that of the German Neuropathic Pain Network.[31] Participants judged the pain intensity evoked by suprathreshold stimuli, verbally reporting a number between 0–100, without a visual reference. Participants were instructed that "0" represented no pain, while "100" represented the most intense pain imaginable. Participants reported pain intensity after a single stimulus (applied for approximately 0.5 sec), and then again after a series of 10 stimuli were applied at 1 s intervals. For the series of 10 stimuli, participants were asked to report an overall pain intensity for the series of stimuli. At 15 and 30 sec after the series-of-10 stimuli were administered, participants were asked to rate the pain intensity of any residual sensation at the stimulated finger. Participants were also asked if any residual non-painful sensations were present at the 30s time point. This testing series (a single stimulus followed by a series-of-10) was conducted four times with the 256mN probe, and then with the 512mN probe. TS of pricking pain was calculated as the difference between the rating of the series-of-10 stimuli and the rating of the single stimulus. For these suprathreshold rating protocols, participants were first given "practice runs" on a site distant from subsequent testing, in order to verify the participant's understanding of the protocol. In the course of this testing, if the participant reported "100," the protocol was halted. The participant was then offered the option of continuing with the next series, or omitting the rest of this set of tests. The participant was also instructed that s/he could stop testing at any time by telling the examiner to stop.

Heat pain—Heat pain sensitivity was assessed using a commercially available thermal stimulator (Pathway; Medoc; Ramat Yishai, Israel). Stimuli were applied on the ventral forearm. Heat pain threshold was determined using a protocol similar to that for PPT. The ATS thermode ($2.56cm^2$) was manually placed in contact with the skin at a temperature of 32°C. After a few seconds, the temperature increased at a rate of 0.5°C/s until the participant pushed a button indicating s/he just then felt a pain sensation. The temperature of the

thermode at the time of the button press was recorded as a threshold estimate. This was repeated four times, moving the thermode to a new site on the forearm each time. Following this, pain tolerance was estimated using the same protocol. The sole difference was that the participant was instructed to press the button when s/he could no longer tolerate the pain. This was repeated four times, moving the thermode for each trial. For both threshold and tolerance testing, a ceiling temperature was set at 52°C, which was entered as the threshold or tolerance estimate if the participant failed to press the button on a given trial. For both the threshold and tolerance protocols, participants were first given "practice runs" on a site distant from subsequent testing, in order to verify the participant's understanding of the protocol.

Following heat pain tolerance testing, participants judged the pain intensity evoked by suprathreshold heat stimuli, verbally reporting a number between 0–100. As with pricking pain ratings, participants were instructed that "0" represented no pain, while "100" represented the most intense pain imaginable. Participants were told that they would receive 10 thermal stimuli in a row, and would be verbally cued to report their peak pain intensity after each stimulus. Practice trials were administered to provide the participant a sense of the timing of stimulus delivery, and to verify understanding of the protocol. The CHEPS thermode ($5.73cm^2$) was manually placed on the skin at a temperature of 38°C, and then a series of 10 temperature pulses were given at 2.4–2.5 sec interstimulus interval. For the first series of stimuli, the peak temperature was 46°C, with a ramp rate of 20°C/s, and a hold time of 750msec at the peak temperature. The participant was cued to report when the temperature just started to decline after reaching the peak. At intervals of 15 and 30 seconds after the 10th thermal pulse, the participant was asked to rate the pain intensity of any lingering sensation using the same 0–100 scale. Following this, the thermode was moved to another location on the forearm, and the same protocol was conducted with a peak temperature of 48°C. This was followed by another test series with a 50°C peak temperature at a third location. In the course of the testing, if the participant reported "100", the thermode was removed and that series of thermal stimuli was halted. The participant was then offered the option of continuing with the next series, or omitting the rest of this set of tests. The participant was also instructed that s/he could stop testing at any time by telling the examiner to stop.

An inadvertent protocol variation at one of the study sites – using a different baseline temperature (38°C) for heat pain threshold and tolerance protocols – led to separate evaluation of these data. Comparison of threshold data from the one site vs. the others revealed a clear difference in the distributions, indicating that the values from the one site should be excluded from analysis. Comparison of tolerance data showed very little difference among the sites, indicating that these data could be combined without a bias imposed by the procedural variation.

Statistical methods

Missing Data Imputation—Imputation for missing values used an expectation-maximization method (as described in [34]) which finds maximum likelihood estimates using a parametric model for incomplete data. For most summary measures, the criteria for imputation on QST items were based on having at least 50% of the data for any type of measure for a given participant. Specifically, data were imputed for individuals with at least 8 valid values among the 16 measurements from the 256 mN and 512 mN probes on single stimulus and series of ten stimulus measurements, whereas data are not imputed for individuals with fewer than 8 valid values. Similarly, individuals with at least 8 of the 16 aftersensation measurements at 15 and 30 seconds for 4 trials each were included in imputation for post-stimulus ratings. For PPT data, imputation was conducted for

individuals with at least 10 of the 20 measures on the 5 body sites: temporalis, masseter, TMJ, trapezius, and lateral epicondyle, each measured twice bilaterally. For thermal measures, data were imputed for individuals with at least 4 of the combined 8 threshold and tolerance ratings from temperatures 1 to 4. Similarly, imputation was done for individuals with at least 3 of the 6 aftersensation measurements at both 15 and 30 seconds from temperatures 46, 48, and 50°C. An exception to this 50% criterion was made for thermal ratings, whereby data were imputed for individuals with at least one of the 30 measurements from the thermal trains at 46, 48, and 50°C. The rationale for this criterion stems from the data collection protocol which required a halt in the procedure when participants gave a rating of 100, with the option to continue at subsequent temperatures. Thus an individual with only one rating at 46°C is informative, and their exclusion would certainly bias results. The statistics regarding data imputation for each variable are presented in Supplementary e-Tables 1 and 2.

Group Comparisons—Descriptive statistics for each summary score were generated. Statistical significance of differences in mean scores was evaluated using an ANOVA derived from a least squares general linear model in which study site was a covariate. Study site was used as a covariate because operational requirements during recruitment created different proportions of cases among sites. The relationship between each summary score and occurrence of TMD was expressed as the standardized odds ratio (SOR), calculated from an unconditional, binary logistic regression model. To achieve this, the summary score was transformed to a unit-normal deviate, and study site was a covariate. The transformation meant that odds ratios could be interpreted as the relative change in odds of TMD for each standard deviation of change in the summary score. In order to be able to directly compare SORs derived from threshold/tolerance measures with all the others, the inverse of threshold/tolerance SORs are presented, so that for all measures, a higher value represents greater pain sensitivity for TMD patients vs. controls. A second logistic regression model generated a fully-adjusted estimate of the relationship, using additional covariates of age (in years), gender, and race/ethnicity (dichotomized as white or non-white). A third logistic regression model using the imputed dataset calculated standardized odds ratios for each summary score, with adjustment for study site, age, gender, and race/ethnicity.

All P-values were computed without adjustment for multiple tests, and we therefore refrain from nominating P=0.05 as a threshold for statistical significance. In this paper's case-control analysis, 40 measures were investigated and, therefore, Bonferroni correction for the probability of type I error would yield a critical P value of 0.05 / 40 = 0.00125. Using the same rationale, rejection of the null hypothesis concerning odds ratios would occur only if the 99.875% confidence interval excluded the null value of one. In general, though, we avoid drawing conclusions about statistical significance of associations, even with correction for multiple tests, because these papers report only univariate- or demographically-adjusted results. Furthermore, the Bonferroni adjustment is probably overly-conservative in this setting, where several measures are moderately correlated. Instead, we will reserve judgments about statistical significance to subsequent papers that will use multivariable modeling to consider multiple characteristics simultaneously, as proposed in the OPPERA heuristic model.

Of the 185 TMD cases, only 19 had arthralgia with no myalgia, and only 9 had myalgia with no arthralgia (the remaining 157 cases were classified with both myalgia and arthralgia). The small numbers with myalgia alone or with arthralgia alone precluded any useful analysis of those TMD subtypes.

Principal Component Analysis—Additionally, we applied principal component analysis (PCA) to this data set to reduce the dimensionality and to identify putative latent

variables. The approach began with four steps, widely used in exploratory principal component analysis:[27] (1) variable selection; (2) evaluation of the correlation matrix; (3) extraction of principal components; and (4) varimax rotation and interpretation of factor loadings. This model is being used for purely exploratory purposes. In particular, this model will not be used to calculate summary scores for the various components.

Our primary interest focused on PCA loadings in controls for three reasons: (a) there are many more controls than cases, thereby improving statistical power to identify factors and estimate loadings; (b) our underlying conceptual model of TMD[22] proposes that relationships between putative risk factors might alter following onset of chronic pain, and we wanted to evaluate that possibility; (c) the long-term goal of OPPERA is to identify risk factors for TMD among controls, so it is desirable to identify a reduced set of variables for that group. Thus, we fit separate PCA models for the TMD cases and controls.

We included the variables identified in Tables 1–2 in the model, with two exceptions related to cutaneous mechanical pain, and one exception related to heat pain. The "Overall ratings of 10 stimuli"—one for each of the two probe intensities—were excluded since they were derivative of the TS measures, which were considered more important. Given that heat pain thresholds from one of the sites was excluded from analysis, we elected to remove that variable from the PCA. This allowed us to include all the other data from all four sites. We elected to keep all of the remaining 33 variables in the model even though some of the variables were highly correlated with one another. We wanted to determine which measures could be combined into latent variables, and this required that all the variables be retained in the model. Although this has the potential to increase the variance of our estimates of the PCA loadings, this was not a major concern given our very large sample size. As shown below, we estimated the variance of our PCA loadings using bootstrapping, and these variances were uniformly low despite the correlations among the input variables. Also, since we are not creating summary scores, there is no conceptual problem with having multiple items that measure similar constructs loading on a single component.

After imputing missing values as described earlier (see preceding paragraph), there were 291 participants who still had at least one missing value among the variables that we included in the PCA model. Moreover, some study sites had more participants with missing values than others. Given the danger of bias associated with dropping 291 participants from the model, we performed a second round of imputation. First, we imputed any missing observations of mechanical cutaneous pain variables based on the observed mechanical cutaneous pain variables for each participant. We also imputed missing heat pain variables based on the observed heat pain variables. We used the EM algorithm to perform this imputation under the assumption that the data are multivariate normal, as described previously. Finally, we imputed any remaining observations of any of the variables in the model based on all the observed QST variables for that participant across all domains, namely: pressure pain, cutaneous mechanical pain, and heat pain. After performing this second round of imputation, there were no remaining participants with missing data. PCA was performed on this fully imputed data set.

A scree plot was generated to estimate the number of components to include in the model (Figure 1). The variance explained by each principal component decreased most conspicuously after the fifth component, suggesting that at least five components are needed. This estimate was verified with a parallel analysis[38]. Parallel analysis estimates the number of components to include in a PCA model by generating random data sets with the same numbers of observations and predictor variables as the original data. Eigenvalues were computed for each random data set and averaged over all the data sets. When the average eigenvalue from these randomly generated data sets is larger than the corresponding

eigenvalue of the original data, then the principal component associated with that eigenvalue is likely to be random noise. The parallel analysis also showed strong evidence that components one through five were above the chance line (Figure 1). Components six and seven were near the chance line, so it was unclear if they should be included in the model.

We therefore fit PCA models using five, six, and seven components. The bootstrap confidence intervals for components six and seven were very wide for many of the loadings, suggesting that the estimated loadings were unstable. (Data not shown) However, the confidence intervals for the loadings of the first five components were very narrow, indicating a stable (and accurate) model. Additionally, the Cronbach's alpha values for these five components were high—ranging from 0.82–0.94 — further supporting a good reliability of this model. Thus, we report a model based on five components.

We fit the PCA models using the R statistical computing platform. The models for the TMD cases and controls were fitted separately. All variables were normalized to have mean 0 and standard deviation 1 prior to fitting the models. After calculating the PCA eigenvectors, we applied a varimax rotation to increase the interpretability of the resulting PCA loadings. (Non-orthogonal rotations were also considered, but the resulting loadings were nearly identical to the loadings produced by the varimax rotation.) Subsequent descriptions of the PCA loadings refer to the rotated loadings.

We estimated the variance in the PCA loadings of each model by drawing 1,000 bootstrap samples for each data set and fitting a PCA model for each replicate. The 95% confidence bounds for our PCA loadings were estimated to be the 2.5% and 97.5% percentiles of the corresponding loading over the 1000 bootstrap replicates [10].

A second PCA was conducted to determine the place of heat pain thresholds in the component weights. In this case, data were derived from only the three study sites employing the same testing protocol.

Much of the data from this study is elaborated in Supplementary e-Tables. E-Tables 1–2 enumerate missing and imputed data for each QST measure. E-Tables 3–5 report means and distributions for each QST measure. E-Tables 6–9 elaborate the results of the PCA analyses for both TMD cases and controls. E-Tables 10–15 report the gender, race/ethnicity, and age effects among the control group's QST measures.

RESULTS

Case-control comparisons

All P-values were computed without adjustment for multiple tests. A decision was made not to use Bonferroni (or other) methods to correct for multiple comparisons. We avoid drawing firm conclusions about statistical significance of associations, even with correction for multiple tests, as this paper reports only univariate- or demographically-adjusted results. Instead, exacting judgments about statistical significance will await subsequent papers that will use multivariable modeling to consider multiple characteristics simultaneously.

For nearly all measures, cases were significantly more sensitive, on average, than controls (Tables 1–2; Supplementary e-Tables 3–5). At all bodily sites, pressure pain thresholds (PPTs) were lower in cases than controls. Exceptions were found for heat pain TS measures, which failed to show statistically significant case-control differences for all six measures. Additionally, the ratings of individual heat pain stimuli at 48°and 50°C failed to show statistically significant case-control differences, although the same trend was apparent as the significant difference found for ratings of 46°C. Consideration of the individual trial ratings

reveal that for all trials, cases rated the heat pain as more intense than controls, while the increase in ratings with repetition was only slightly greater for cases than controls (Figure 2).

Standardized odds ratios (SORs; inverted for threshold/tolerance measures) were above 1.0, signifying that greater odds of TMD were associated with greater ratings of pain or lower threshold/tolerance. In all instances, with the exception of the heat pain TS measures, the ratings of 48°C stimuli, and heat pain threshold (with imputed values), the 95% CI of the SORs excluded the value of 1.0, reflecting a statistically significant difference between the groups. Among the threshold and tolerance measures, SORs for the PPTs were the highest (ranging from 2.38–4.05), particularly for the cranial sites. Furthermore, the 95% CIs for SORs of trapezius and cranial site PPTs are distinctly outside the range of SORs for any of the rating measures, indicating a statistically significant difference between the PPTs and any of the rating measures. The SOR for mechanical cutaneous pain threshold (2.01) was similar to that for PPT of the lateral epicondyle (2.38), with largely overlapping 95% CI's. In contrast, the 95% CIs for both of these measures did not overlap with that of heat pain threshold, indicating a statistically significant difference in SORs for mechanical pain vs. heat pain thresholds. Among all the suprathreshold rating measures (ratings of single stimuli, aftersensations, and TS; ranging from 1.07–1.44), 95% CI's overlapped considerably, suggesting no statistically significant difference among any of these measures in their ability to distinguish TMD cases from controls.

The SORs for all QST measures changed little when adjusted for multiple factors, or comparing results with and without missing data imputation (Tables 1–2).

PCA

The loadings (Table 3) and corresponding 95% confidence intervals (Supplementary e-Table 6) show the PCA results derived from the control participants. Component 1 is primarily based on the heat pain ratings of suprathreshold stimuli, with a weaker loading from heat pain tolerance. Component 2 is primarily based on the heat pain aftersensation ratings, with weaker loadings from the mechanical pain aftersensation ratings derived from the larger intensity stimulus. Component 3 is primarily based on the mechanical cutaneous pain measures: ratings, TS, aftersensations, and threshold. Component 4 is principally based on pressure pain thresholds. Component 5 is primarily based on heat pain TS. Of note, the mechanical TS derived from the higher intensity stimulus did not load above 0.40 on any of the factors.

A second PCA was conducted on data from control subjects, this time including heat pain threshold data, which restricted the PCA to data derived from the three sites using the same heat pain threshold parameters. A similar set of components were produced, with the highly weighted variables still defining the five components as described above. Heat pain threshold had the strongest weighting in Component 4 (0.49), and weightings below 0.40 for other components. Thus, it more strongly associated with pressure pain thresholds than with any of the other heat pain measures.

PCA results for the TMD cases (Supplementary e-Tables 7–8) show a similar grouping, with small differences. For one thing, the ordering of the components (based on percent of variance accounted for) is different, although this reflects very small differences in the actual variance values. With respect to individual loadings, the mechanical pain aftersensation ratings load strongly with both components 1 and 5 for the TMD cases (corresponding to components 2 and 3 for the controls). In this manner, it "split" its association between heat pain aftersensation and other mechanical pain measures. In contrast, these measures for the control group only significantly associated with other

mechanical pain measures. Additionally, heat pain intensity ratings load strongly on component 2 (component 1 for controls), but also load negatively on component 3 (component 5 for controls), which otherwise has strong positive loadings from heat pain TS measures. This negative loading was indicated, but of insufficient magnitude to be deemed significant in the control group. Given these apparent differences, a series of permutation tests were performed to test the null hypothesis that the loadings were the same for both models (Supplementary e-Table 9; component numbering matching that for controls). Of the differences noted above, the loadings of heat pain single stimulus ratings upon TMD component 3 (control component 5) were statistically significantly different between TMD cases and controls at the p=0.05 level. However, any adjustment for multiple testing would render these differences non-significant. Otherwise, a scattering of other loading differences were statistically significant at p=0.05, but of low loading overall. The corresponding confidence intervals for the model based on TMD cases are slightly wider (Supplementary e-Table 6 vs. Supplementary e-Table 8), but this is not surprising, given that the sample size was much smaller. Neither model shows evidence of high variance or other forms of instability.

Demographic differences

Among controls, women showed greater pain sensitivity than men for almost all pain measures (Supplementary e-Tables 10–11). The exceptions were the heat pain TS measures, which were not significantly different between the sexes.

Several pain measures revealed less pain sensitivity for non-Hispanic whites compared to other racial/ethnic groups combined (Supplementary e-Tables 12–13). Exceptions to this included the cranial pressure pain thresholds, mechanical cutaneous and heat pain thresholds, and some of the heat pain TS measures, which did not show significant group differences.

A number of pain measures differed significantly among age groups (Supplementary e-Tables 14–15). In most instances, differences were in the direction of greater pain sensitivity for the younger participants. This was observed for some of the pressure pain thresholds (marginally), all of the mechanical cutaneous pain measures, and some of the heat pain measures, most strongly the aftersensation ratings.

DISCUSSION

The key findings from this study are as follows:

1. Chronic TMD cases were more pain sensitive than controls to a wide range of mechanical and thermal stimuli, applied to symptomatic and asymptomatic body sites. This difference was most pronounced for pressure pain thresholds, and less pronounced for most of the cutaneous mechanical and heat pain sensitivity measures. One aspects of heat pain sensitivity, temporal summation, did not distinguish TMD cases from controls.

2. Mechanical and heat pain sensitivity measures can be regarded as reflecting five components of experimental pain sensitivity: (1) heat pain ratings, (2) heat pain aftersensations and tolerance, (3) mechanical cutaneous pain sensitivity, (4) pressure pain thresholds, and (5) heat pain temporal summation. These components were similar for TMD cases and non-TMD controls.

3. Demographic differences were found for multiple experimental pain measures, most strongly and widely for sex differences, somewhat less so for racial/ethnicity differences, and for age differences.

TMD case-control differences in experimental pain sensitivity

A number of studies have reported TMD case-control differences in experimental pain sensitivity, typically using one or two pain measures,[23–25,32,37] although not universally.[3,37] In all instances of significant differences, chronic TMD cases were found to be more pain sensitive. This current study supports these previous reports, with a much larger sample size, and also provides the opportunity to compare pain sensitivity differences across a large array of tests. With respect to PPTs, the statistical tests and the (reverse coded) SORs would be expected to be highly significant at cranial sites, given that these sites are symptomatic regions for TMD. Yet, the (reverse coded) SORs for the PPTs on extra-cranial sites are also highly significant, indicating that remote muscle hyperalgesia is prominent in TMD cases. The cutaneous mechanical pain threshold (reverse coded) SOR, while also highly significant, was smaller and had 95% CIs that did not overlap with cranial site PPTs, thus making it less discriminating of case-control differences than the PPTs. Furthermore, the heat pain threshold (reverse coded) SOR, while also significant, was smaller still and had 95% CIs that did not overlap with any of the other threshold's (reverse coded) SORs, making it least discriminating of case-control differences among all the threshold measures. It should be noted that the mechanical cutaneous pain and heat pain testing only involved asymptomatic body sites (the upper extremity). As such, they represent an index of remote or widespread hyperalgesia for the TMD cases.

With respect to suprathreshold ratings of single stimuli, the SORs for mechanical cutaneous stimuli were higher than those for heat stimuli, showing the same trend revealed by thresholds. In fact, while all the mechanical cutaneous measures showed significant case-control differences, not all the heat pain measures did. Specifically, the fully adjusted heat pain TS SORs showed the same trend in case-control differences, but did not reach statistical significance. However, the fact that 95% CIs for most of these SORs overlap mitigate against the general hypothesis that suprathreshold cutaneous mechanical pain measures discriminate TMD cases from controls better than suprathreshold heat pain measures do.

SORs for ratings of aftersensation pain intensity showed similar magnitudes of association with TMD as the single trial ratings. This observation suggests that exaggerated sensory perseveration is as prominent as cutaneous hyperalgesia in TMD cases.

The heat pain TS measures—whether calculated as a difference in ratings between the first and later stimuli, or as the rate of change in pain ratings over the first few stimuli in a series —did not show statistically significant case-control differences. This contrasts with earlier reports of increased heat pain TS in women with TMD [25]. It is worth noting that the curves depicting TS in Figure 2 of this paper—showing the higher ratings of pain for TMD cases across all trials—looks very similar to Figure 4B of Maixner et al.[25]. In this respect, these results are consistent. Also, TS can be calculated in multiple ways, and is complicated by ceiling effects of a bounded rating scale, as used here. The current report employed two commonly used measures for heat pain TS, and used a conservative process to adjust for ceiling effects. A subsequent report will examine these data in a more elaborate way, in order to better account for individual variation in TS profiles and ceiling effects.

It is worth noting that our results are consistent with a recent report, in which TMD cases were not significantly different from controls in measures of heat pain TS, but showed significantly higher heat pain aftersensation ratings.[29]

Mechanical pain TS was significantly higher for TMD cases than controls, yet heat pain TS measures did not differ significantly between the two groups. This may be due to mechanistic differences between mechanical and thermal TS. However, protocol differences

may also be responsible for differences in our TS results. For heat pain, individual stimuli were rated, and TS measures were calculated by virtue of rating changes from one stimulus to the next. For mechanical pain, a rating was given for a single stimulus, and then another rating was given for a series of ten stimuli applied at 1 Hz, and the difference between these two ratings were taken as the TS. The extent to which these methodological differences could contribute to different results cannot be separated from potential mechanistic differences.

In comparing SORs from site-adjusted vs. fully-adjusted analyses, the latter typically had slightly higher values. Thus, the case-control differences were not seriously confounded by demographic differences between TMD cases and controls. There were few or no changes in fully-adjusted SORs when comparing analyses conducted with unimputed vs. imputed data. Thus, the process of data imputation performed for this study, while potentially changing the results by way of reducing the bias of data dropout, had no substantive effect upon the outcome.

PCA-derived components of experimental pain sensitivity

PCA analysis on control participants yielded a 5-component solution that was parsimonious and robust. The components were characterized as "heat pain ratings" (component 1), "heat pain aftersensations and tolerance" (component 2), "mechanical cutaneous pain sensitivity" (component 3), "pressure pain thresholds" (component 4), and "heat pain TS" (component 5). Of note, all components are modality specific.

Modality specific components have been described in a previous report of factor analyses of experimental pain measures.[16] Another study evaluating associations among multiple experimental pain measures found higher associations within than across modalities.[2] In contrast to these studies, an analysis of pain threshold data across multiple stimulus modalities found only a small fraction of the variance attributable to stimulus modality.[26] The results reported here support the previous studies that show experimental pain measurement factors or components to be modality specific.

It is reasonable to expect that the experimental pain measures may separate according to stimulus modality. The nervous system components that relay nociceptive information show some degree of modality specificity. For instance, along with polymodal nociceptors, which respond to both thermal and mechanical stimuli, there are also nociceptors that only respond to mechanical stimuli and others that only respond to heat.[13,20,35] Furthermore, while many spinal cord and thalamic neurons respond to both noxious thermal and mechanical stimuli, there are those that only respond to noxious mechanical stimuli.[7,19] Thus, at the individual neuron level, nociceptive information initiated by heat vs. mechanical stimuli is partially segregated in the CNS. This separation of neural processing would be expected to play a role in perceptual differences, and result in increased inter-modality vs. intra-modality variance.

Heat pain TS measures did not load with heat pain ratings on a common component, in agreement with a previous study.[16] This finding suggests that TS measures reflect mechanisms that are distinct from those responsible for simple suprathreshold heat pain perception. Indeed, some differences, such as the key role of the *N*-methyl *D*-aspartate (NMDA) receptor in heat pain TS, have been documented.[39]

The current study is the first to report data on both mechanical and thermal TS obtained from the same participants. Strikingly, the two measures did not load together in the PCA analyses, suggesting significant differences in the mechanisms underlying TS between the

two modalities. However, as noted above, methodological differences may have played a role in differentiating these measures.

Demographic Differences

In accordance with many previous reports, women showed significantly greater pain sensitivity than men in almost all of the QST measures used in this study. A recent review[11] tabulated results from published studies of this topic, and found that 68 −75% of studies reported significantly lower thresholds or significantly greater TS for women when tested with modalities used in the present study. The one exception to finding sex differences in the present study was with respect to TS of heat pain. Of four published studies evaluating sex differences in heat pain TS, three reported significantly greater TS in women,[12,14,30] and one reported no difference.[36] Thus, a large majority of laboratory pain studies investigating sex differences found greater sensitivity among women than men, although the magnitude of the difference varies across studies and across pain measures.

Most pain measures revealed significantly less pain sensitivity for non-Hispanic whites than non-white races and Hispanics, combined. This was true for all the mechanical cutaneous pain measures except threshold, and for the heat pain measures, except for threshold and some of the TS measures. Interestingly, the pressure pain thresholds showed significant differences on the trapezius and lateral epicondyle sites, but not any of the cranial sites. These results are somewhat consistent with previous reports of greater pain sensitivity in African-American participants vs. non-Hispanic while participants, particularly for suprathreshold pain assessments.[4,28]

Fewer pain measures showed significant age effects. In most cases, differences were in the direction of greater pain sensitivity for the younger participants. In comparison to sex and racial differences in pain measures, age differences showed smaller mean value differences. It should be noted that the age range of participants for this study (18–44) is not large compared to studies specifically evaluating the effects of age upon pain. Most studies have reported no significant age effects upon experimental pain within this age range, but only when participants in this age range were compared with more elderly participants.[15]

Study limitations

One should note that the large sample size for this study provides for significant group differences even with small mean differences. As one example, case-control differences in the pain intensity ratings of cutaneous mechanical probes were statistically significant, even though the mean rating difference was less than 10 on the 0–100 scale (Table 1). This raises a cautionary note of distinguishing between statistical significant differences and clinically meaningful differences in pain sensitivity. At the same time, such results exemplify that QST measures can reveal group differences in pain sensitivity that are of a magnitude likely to be missed in a clinical exam.

As is the case for many other studies of demographic differences in experimental pain, several of the observed differences in the present study were either unadjusted, or were adjusted only for study site. Importantly, there was no adjustment for etiological characteristics (e.g., psychological states or other demographic variables) that might confound these demographic associations. Hence, while these observed differences are useful to illustrate demographic risk indicators and potential sources of confounding in our case-control comparisons, they do not necessarily signify causal effects of demographic characteristics.

We formally assessed inter-examiner reliability of PPT, as this QST measure was the most likely to be variably influenced by the examiner's handling of the stimulus probe. We found

a very good level of reliability (overall ICC of 0.91), giving us confidence in consistent assessments across examiners and sites. A recent study reported inter-examiner ICCs for several different sensory testing protocols, including some closely approximating those used here (Pigg et al., 2010). They also report a high ICC for PPT (0.89), and nearly as high value for heat pain threshold (0.87). Somewhat lower values were found for mechanical cuntaneous threshold (0.56) and mechanical cutaneous temporal summation (0.52).

Conclusions

This large scale study demonstrated significantly greater pain sensitivity for people with TMD vs. those without TMD in a wide range of mechanical and thermal pain tests. Most of the tests involved stimulation of asymptomatic body sites on the upper extremity, providing support for the theory of a generalized upregulation of pain processing in chronic pain conditions, even those with somatically restricted symptomology.

PERSPECTIVE

This article describes experimental pain sensitivity differences between a large sample of people with chronic TMD and non-TMD controls, using multiple stimulus modalities and measures. Variability in the magnitude and consistency of case-control differences highlight the need to consider multiple testing measures to adequately assess pain processing alterations in chronic pain conditions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Reference List

1. Arendt-Nielsen, L.; Lautenbacher, S. Assessment of pain perception. In: Lautenbacher, S.; Fillingim, RB., editors. Pathophysiology of Pain Perception. New York: Kluwer Academic; 2004. p. 25-42.

2. Bhalang K, Sigurdsson A, Slade GD, Maixner W. Associations among four modalities of experimental pain in women. J Pain. 2005; 6:604–611. [PubMed: 16139779]

3. Bragdon EE, Light KC, Costello N, Sigurdsson A, Bunting S, Bhalang K, Maixner W. Group differences in pain modulation: pain-free women compared to pain-free men and to women with TMD. Pain. 2002; 96:227–237. [PubMed: 11972994]

4. Campbell CM, Edwards RR, Fillingim RB. Ethnic differences in responses to multiple experimental pain stimuli. Pain. 2005; 113:20–26. [PubMed: 15621360]

5. Carlson CR, Reid KI, Curran SL, Studts J, Okeson JP, Falace D, Nitz A, Bertrand PM. Psychological and physiological parameters of masticatory muscle pain. Pain. 1998; 76:297–307. [PubMed: 9718248]

6. Cornsweet TN. The staircase-method in psychophysics. American Journal of Psychology. 1962; 75:485–491. [PubMed: 13881416]

7. Davidson S, Zhang X, Khasabov SG, Simone DA, Giesler GJ Jr. Termination Zones of Functionally Characterized Spinothalamic Tract Neurons Within the Primate Posterior Thalamus. Journal of Neurophysiology. 2008; 100:2026–2037. [PubMed: 18701750]

8. Diatchenko L, Nackley AG, Slade GD, Fillingim RB, Maixner W. Idiopathic pain disorders - Pathways of vulnerability. Pain. 2006; 123:226–230. [PubMed: 16777329]

9. Dworkin SF, LeResche L. Research diagnostic criteria for temporomandibular disorders: Review, criteria, examinations and specifications, critique. J Craniomandib Disord. 1992; 6:301–355. [PubMed: 1298767]

10. Efron, B.; Tibshirani, R. An Introduction to the Bootstrap. Boca Raton, FL: Chapman & Hall/CRC; 1994.

11. Fillingim RB, King CD, Ribeiro-Dasilva MC, Rahim-Williams B, Riley JL III. Sex, Gender, Pain: A Review of Recent Clinical and Experimental Findings. Journal of Pain. 2009; 10:447–485. [PubMed: 19411059]

12. Fillingim RB, Maixner W, Kincaid S, Silva S. Sex differences in temporal summation but not sensory- discriminative processing of thermal pain. Pain. 1998; 75:121–127. [PubMed: 9539681]

13. Garell PC, McGillis SLB, Greenspan JD. Mechanical response properties of nociceptors innervating feline hairy skin. Journal of Neurophysiology. 1996; 75:1177–1189. [PubMed: 8867127]

14. George SZ, Wittmer VT, Fillingim RB, Robinson ME. Sex and Pain-Related Psychological Variables Are Associated With Thermal Pain Sensitivity for Patients With Chronic Low Back Pain. The Journal of Pain. 2007; 8:2–10. [PubMed: 17207739]

15. Gibson SJ, Farrell M. A review of age differences in the neurophysiology of nociception and the perceptual experience of pain. Clin J Pain. 2004; 20:227–239. [PubMed: 15218407]

16. Hastie BA, Riley JL III, Robinson ME, Glover T, Campbell CM, Staud R, Fillingim RB. Cluster analysis of multiple experimental pain modalities. Pain. 2005; 116:227–237. [PubMed: 15964682]

17. Janal MN, Glusman M, Kuhl JP, Clark WC. On the absence of correlation between responses to noxious heat, cold, electrical and ischemic stimulation. Pain. 1994; 58:403–411. [PubMed: 7838590]

18. Lautenbacher S, Rollman GB, McCain GA. Multi-method assessment of experimental and clinical pain in patients with fibromyalgia. Pain. 1994; 59:45–53. [PubMed: 7854801]

19. Lenz FA, Gracely RH, Rowland LH, Dougherty PM. A population of cells in the human thalamic principal sensory nucleus respond to painful mechanical stimuli. Neuroscience Letters. 1994; 180:46–50. [PubMed: 7877759]

20. Lynn B, Faulstroh K, Pierau F-K. The classification and properties of nociceptive afferent units from the skin of the anaesthetized pig. European Journal of Neuroscience. 1995; 7:431–437. [PubMed: 7773440]

21. Lynn B, Perl ER. A comparison of four tests for assessing the pain sensitivity of different subjects and test areas. Pain. 1977; 3:353–365. [PubMed: 909713]

22. Maixner W, Diatchenko L, Dubner R, Fillingim RB, Greenspan JD, Knott C, Ohrbach R, Weir B, Slade GD. Orofacial Pain: Prosepctive Evaluation Risk Assessment - The OPPERA study. Journal of Pain. 2011 this issue.

23. Maixner W, Fillingim R, Booker D, Sigurdsson A. Sensitivity of patients with painful temporomandibular disorders to experimentally evoked pain. Pain. 1995; 63:341–351. [PubMed: 8719535]

24. Maixner W, Fillingim R, Kincaid S, Sigurdsson A, Harris MB. Relationship between pain sensitivity and resting arterial blood pressure in patients with painful temporomandibular disorders. Psychosomatic Medicine. 1997; 59:503–511. [PubMed: 9316183]

25. Maixner W, Fillingim R, Sigurdsson A, Kincaid S, Silva S. Sensitivity of patients with painful temporomandibular disorders to experimentally evoked pain: evidence for altered temporal summation of pain. Pain. 1998; 76:71–81. [PubMed: 9696460]

26. Neddermeyer TJ, Fluhr K, Lotsch J. Principle components analysis of pain thresholds to thermal, electrical, and mechanical stimuli suggests a predominant common source of variance. Pain. 2008; 138:286–291. [PubMed: 18243556]

27. Pett, MA.; Lackey, NR.; Sullivan, JJ. Making Sense of Factor Analysis: The Use of Factor Analysis for Instrument Development in Health Care Research. Thousand Oaks, CA: Sage Pub.; 2003.

28. Bridgett, Rahim-Williams F.; Riley, Joseph L.; Herrera, Dyanne; Campbell, Claudia M.; Hastie, Barbara A.; Fillingim, Roger B. Ethnic identity predicts experimental pain sensitivity in African Americans and Hispanics. Pain. 2007 May 1; 129(1):177–184. [PubMed: 17296267]

29. Raphael KG, Janal MN, Ananthan S, Cook DB, Staud R. Temporal Summation of Heat Pain in Temporomandibular Disorder Patients. Journal of Orofacial Pain. 2009; 23:54–64. [PubMed: 19264036]

30. Robinson M, Wise E, Gagnon C, Fillingim R, Price D. Influences of gender role and anxiety on sex differences in temporal summation of pain. J Pain. 2004; 5:77–82. [PubMed: 15042515]

31. Rolke R, Baron R, Maier C, Tolle TR, Treede RD, Beyer A, Binder A, Birbaumer N, Birklein F, Botefur IC. Quantitative sensory testing in the German Research Network on Neuropathic Pain (DFNS): Standardized protocol and reference values. Pain. 2006; 123:231–243. [PubMed: 16697110]

32. Sarlani E, Grace EG, Reynolds MA, Greenspan JD. Evidence for up-regulated central nociceptive processing in patients with masticatory myofascial pain. J Orofac Pain. 2004; 18:41–55. [PubMed: 15029872]

33. Sarlani E, Greenspan JD. Evidence for generalized hyperalgesia in temporomandibular disorders patients. Pain. 2003; 102:221–226. [PubMed: 12670662]

34. Slade GD, Bair E, By K, Mulkey F, Baraian C, Gonzalez Y, Gordon S, Ribeiro-Dasilva MC, Lim Pei Feng, Greenspan JD, Dubner R, Fillingim RB, Maixner W, Diatchenko L, Dampier D, Knott C, Ohrbach R. Study methods, recruitment, socio-demongraphic findings and demographic generalizability in the OPPERA study. Journal of Pain. 2011 This issue.

35. Slugg RM, Meyer RA, Campbell JN. Response of cutaneous A- and C-fiber nociceptors in the monkey to controlled-force stimuli. Journal of Neurophysiology. 2000; 83:2179–2191. [PubMed: 10758127]

36. Staud R, Robinson ME, Vierck CJ Jr, Price DD. Diffuse noxious inhibitory controls (DNIC) attenuate temporal summation of second pain in normal males but not in normal females or fibromyalgia patients. Pain. 2003; 101:167–174. [PubMed: 12507711]

37. Svensson P, List T, Hector G. Analysis of stimulus-evoked pain in patients with myofascial temporomandibular pain disorders. Pain. 2001; 92:399–409. [PubMed: 11376913]

38. Thompson B, Daniel LG. Factor analytic evidence for the construct validity of scores: A historical overview and some guidelines. Educational and Psychological Measurement. 1996; 56:197–208.

39. Vierck CJ, Cannon RL, Fry G, Maixner W, Whitsel BL. Characteristics of temporal summation of second pain sensations elicited by brief contact of glabrous skin by a preheated thermode. Journal of Neurophysiology. 1997; 78:992–1002. [PubMed: 9307129]

List of abbreviations

CI	confidence interval
CNS	central nervous system
kPa	kilopascal
kPa/s	kilopascal per second
mN	milliNewton
NMDA	N-methyl-D-aspartate
OPPERA	Orofacial Pain: Prospective Evaluation and Risk Assessment

PCA	Principal Component Analysis
PPT	pressure pain threshold
QST	quantitative sensory testing
SOR	Standardized odds ratio
TMD	Temporomandibular Disorders
TMJ	Temporomandibular Joint
TS	temporal summation
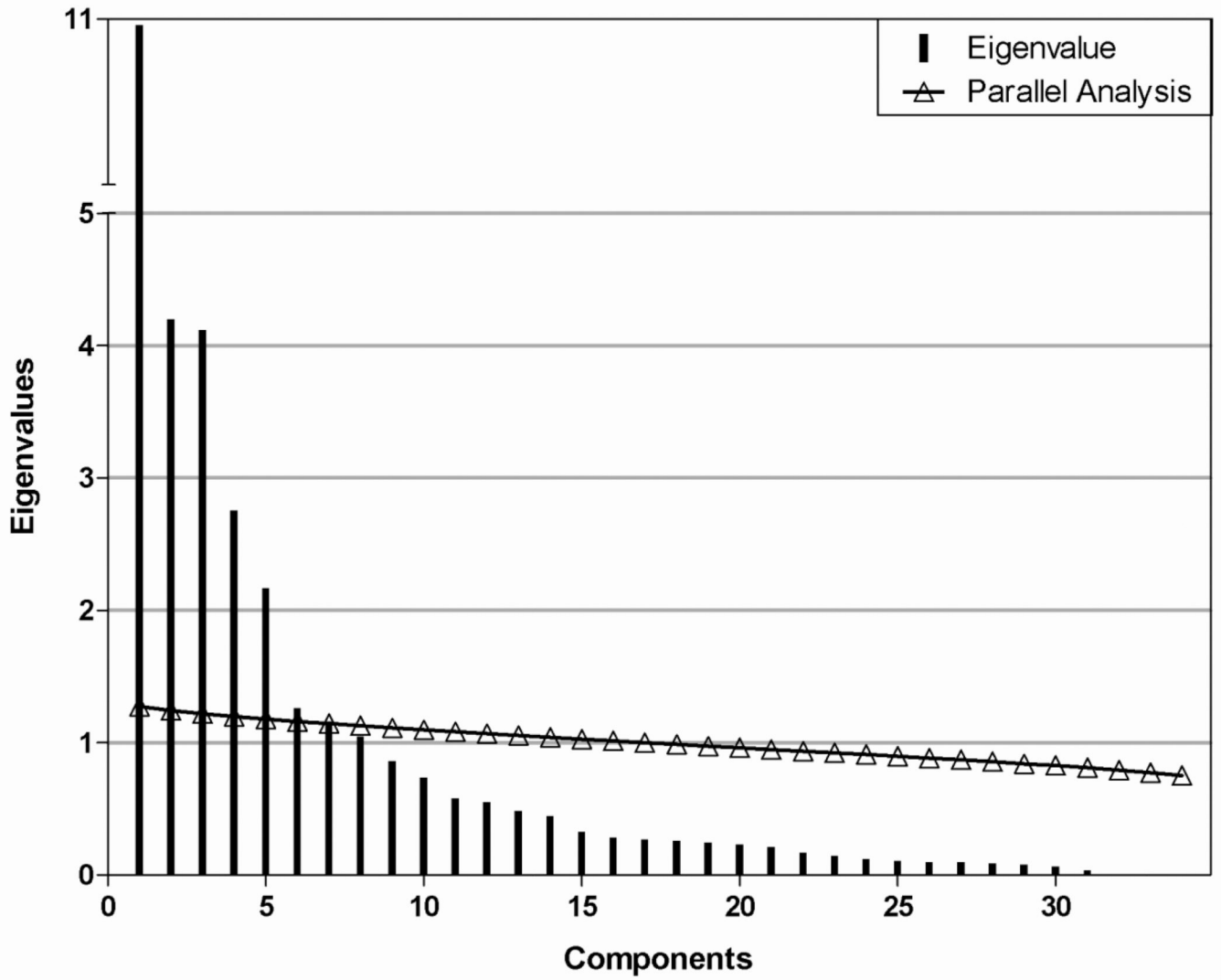
Figure 1.
Components resulting from PCA (vertical bars) and parallel analysis (trianges) conducted on data from control participants.
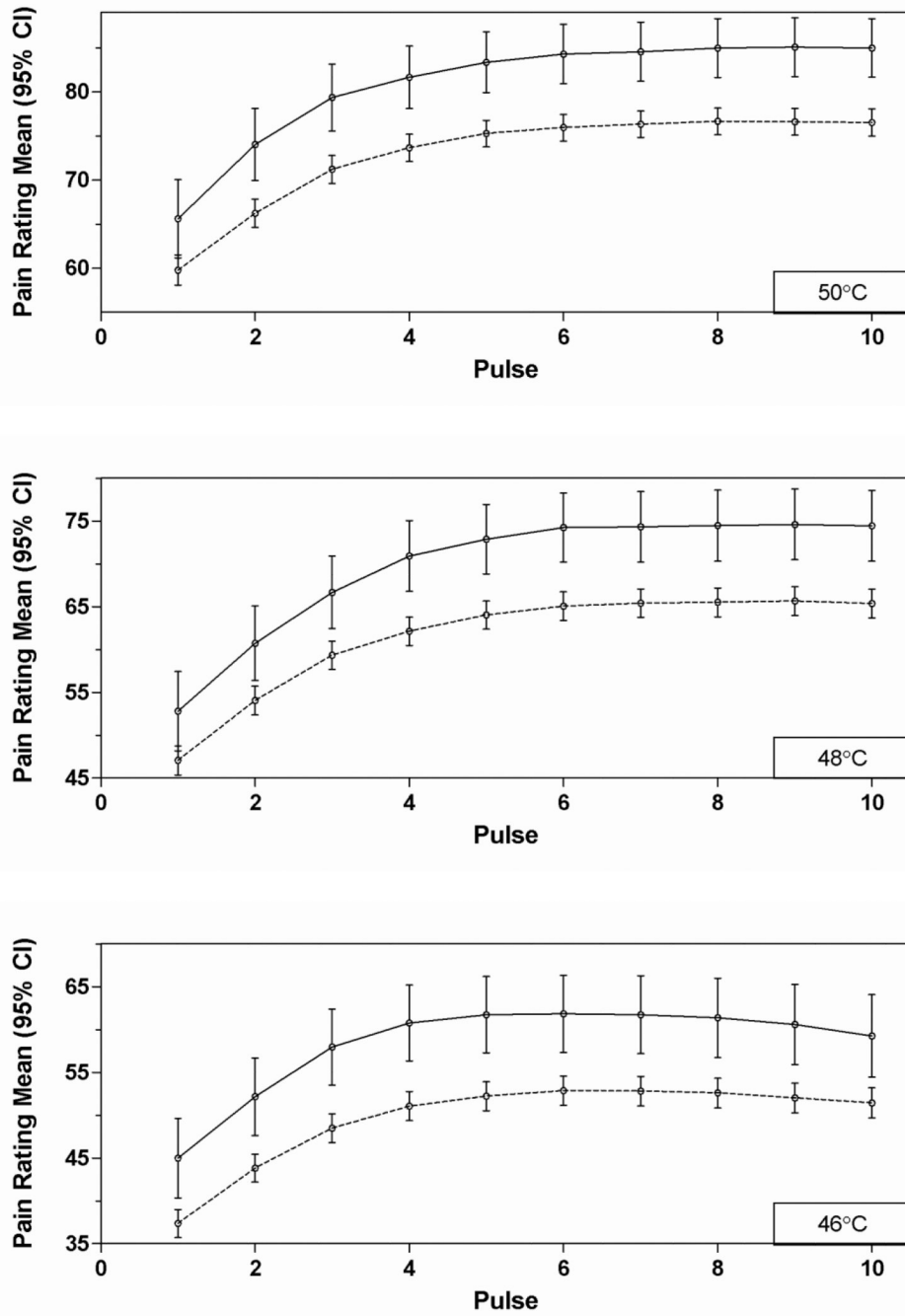
Figure 2.
Ratings of individual heat stimuli (0–100 scale) presented at three different temperatures, ten times each. Solid lines – TMD cases; dashed lines – controls.

Table 1

Variation between TMD cases and controls in mechanical pain measures

Stimulus	Units and potential range	Controls			TMD cases			P-value[1]	Non-imputed effect estimates[2]				Imputed effect estimates		
		n	mean	se	n	mean	se		Site-adjusted effect[3]		Fully adjusted effect[4]		Fully adjusted effect		
									SOR[5]	95%CI[6]	SOR	95%CI	n	SOR	95%CI
Pressure pain threshold[7]															
Temporalis	kPa, 50–600	1601	217.92	2.12	168	148.23	4.06	<0.0001	4.1	3.1,5.4	3.7	2.7,4.9	1793	3.8	2.9,5.1
Masseter	kPa, 50–600	1605	201.01	1.98	160	141.97	4.43	<0.0001	3.4	2.6,4.5	3.0	2.3,4.0	1793	3.4	2.5,4.5
TM joint	kPa, 50–600	1579	184.58	1.72	157	133.84	3.67	<0.0001	3.5	2.7,4.6	3.3	2.5,4.4	1793	3.7	2.8,4.9
Trapezius	kPa, 100–600	1573	372.93	3.70	160	260.71	7.90	<0.0001	2.8	2.2,3.5	2.6	2.0,3.3	1793	2.6	2.1,3.4
Lateral epicondyle	kPa,100–600	1573	389.18	3.69	164	298.88	10.00	<0.0001	2.4	1.9,3.0	2.2	1.7,2.7	1793	2.2	1.8,2.8
Mechanical cutaneous pain															
Threshold[7]	mN, 8–512	1540	269.74	4.30	172	170.45	10.31	<0.0001	2.1	1.7,25	2.0	1.6, 2.5			
Single stimulus ratings															
256mN probe	ratings, 0–100	1562	8.83	0.32	176	13.76	1.20	<0.0001	1.3	1.2,15	1.5	1.3,1.7	1742	1.5	1.3, 1.7
512mN probe	ratings, 0–100	1464	14.79	0.46	161	21.62	1.63	<0.0001	1.3	1.2,15	1.5	1.3, 1.8	1742	1.5	1.3, 1.8
Overall ratings of 10 stimuli															
256mN probe	ratings, 0–100	1563	18.40	0.53	176	28.62	1.82	<0.0001	1.4	1.3,17	1.7	1.4,1.9	1742	1.7	1.4,1.9
512mN probe	ratings, 0–100	1461	31.36	0.71	161	43.69	2.17	<0.0001	1.5	1.3,17	1.7	1.4, 2.0	1742	1.7	1.4, 2.0
Aftersensation ratings															
15 s, 256mN probe	ratings, 0–100	1562	2.90	0.19	174	6.68	0.98	<0.0001	1.3	1.2,15	1.4	1.2,1.6	1742	1.4	1.2,1.6
30 s, 256mN probe	ratings, 0–100	1563	1.35	0.12	175	3.23	0.72	<0.0001	1.2	1.1,14	1.3	1.1, 1.5	1742	1.3	1.1, 1.5
15 s, 512mN probe	ratings, 0–100	1456	6.62	0.33	161	13.75	1.40	<0.0001	1.5	1.3,17	1.6	1.4,1.9	1742	1.5	1.3,1.7
30 s, 512mN probe	ratings, 0–100	1453	3.56	0.22	161	7.16	0.99	<0.0001	1.3	1.2,15	1.4	1.2,1.6	1742	1.3	1.2,1.5
Temporal summation															
256mN probe	Δ ratings, ±100	1560	9.56	0.31	176	14.86	1.05	<0.0001	1.4	1.2,15	1.5	1.3,1.7	1742	1.5	1.3,1.7
512mN probe	Δ ratings, ±100	1461	16.56	0.42	161	22.07	1.26	<0.0001	1.3	1.1,15	1.4	1.2, 1.7	1742	1.4	1.2, 1.6

[1] P-values are from analysis of variance model comparing mean values of putative risk factor between cases and controls, with adjustment for study site.

[2] Non-imputed effect estimates are for complete case analysis using numbers of subjects in columns headed n.

[3] Site-adjusted effects were computed in logistic regression models where the putative risk factor as the main explanatory variable and study site as covariate.

[4] Fully-adjusted effects were computed in logistic regression models that additionally include covariates of age group, gender and race/ethnicity.

[5] SOR is standardized odds ratio from a logistic regression model evaluating the linear effect of the risk factor that was standardized by z-score transformation

[6] 95% confidence interval for SOR

[7] Negative z-scores used for standardized odds ratios which therefore represent increase in odds of TMD associated with reduction of 1 sd in the value of the variable

Table 2

Variation between TMD cases and controls in heat pain measures

Stimulus	Units and potential range	Controls n	Controls mean	Controls se	TMD cases n	TMD cases mean	TMD cases se	P-value[1]	Site-adjusted SOR[5]	Site-adjusted 95% CI[6]	Fully adjusted SOR	Fully adjusted 95% CI	Imputed n	Imputed SOR	Imputed 95% CI
Heat pain threshold[7]	°C, 32–52	1196	42.30	0.09	135	41.44	0.27	0.0027	1.3	1.1, 1.5	1.2	1.0, 1.5	1338	1.2	1.0, 1.5
Heat pain tolerance[7]	°C, 32–52	1506	46.26	0.06	178	45.36	0.18	<0.0001	1.4	1.2, 1.6	1.4	1.2, 1.6	1690	1.4	1.2, 1.6
Single stimulus ratings															
46°C	rating, 0–100	1495	37.34	0.83	174	44.68	2.36	0.0042	1.2	1.0, 1.4	1.3	1.0, 1.5	1672	1.3	1.1, 1.5
48°C	rating, 0–100	1481	46.81	0.86	170	51.72	2.37	0.0664	1.1	0.9, 1.3	1.2	1.0, 1.4	1672	1.2	1.0, 1.4
50°C	rating, 0–100	1430	58.58	0.90	168	64.61	2.31	0.0272	1.2	1.0, 1.4	1.3	1.0, 1.5	1672	1.2	1.0, 1.5
Ratings of 10 stimuli: area under curve															
46°C	ratings, 0–1000	1185	352.73	6.84	132	430.83	18.74	0.0002	1.3	1.1, 1.6	1.4	1.1, 1.7	1672	1.3	1.1, 1.6
48°C	ratings, 0–000	982	416.88	7.72	118	525.77	20.29	<0.0001	1.5	1.2, 1.9	1.6	1.3, 2.0	1672	1.4	1.1, 1.6
50°C	ratings, 0–1000	725	476.73	9.23	74	570.86	24.22	0.0017	1.4	1.1, 1.9	1.6	1.2, 2.1	1672	1.4	1.2, 1.8
Thermal aftersensations															
15 s, 46°C	rating, 0–100	1478	8.49	0.40	169	12.25	1.30	0.0029	1.2	1.0, 1.4	1.2	1.1, 1.4	1653	1.2	1.1, 1.4
30 s, 46°C	rating, 0–100	1479	4.52	0.28	170	6.84	0.91	0.0084	1.2	1.0, 1.3	1.2	1.1, 1.4	1653	1.2	1.1, 1.4
15 s, 48°C	rating, 0–100	1481	12.92	0.49	170	19.09	1.61	<0.0001	1.3	1.1, 1.5	1.5	1.2, 1.7	1653	1.4	1.2, 1.7
30 s, 48°C	rating, 0–100	1483	8.04	0.38	170	11.19	1.28	0.0090	1.2	1.0, 1.3	1.3	1.1, 1.5	1653	1.3	1.1, 1.5
15 s, 50°C	rating, 0–100	1420	14.60	0.53	169	23.70	1.90	<0.0001	1.4	1.3, 1.7	1.6	1.4, 1.9	1653	1.6	1.4, 1.9
30 s, 50°C	rating, 0–100	1418	8.30	0.39	167	13.84	1.53	<0.0001	1.3	1.2, 1.5	1.4	1.2, 1.6	1653	1.4	1.2, 1.6
Temporal summation: highest minus first rating															
46°C	Δ ratings, 0–100	1341	19.82	0.74	157	21.08	2.43	0.5885	1.1	0.9, 1.3	1.0	0.9, 1.2	1672	1.0	0.9, 1.2
48°C	Δ ratings, 0–100	1274	24.13	0.73	151	26.70	2.25	0.2541	1.1	1.0, 1.3	1.1	1.0, 1.4	1672	1.2	1.0, 1.4
50°C	Δ ratings, 0–100	1107	24.50	0.79	131	27.21	2.04	0.2583	1.1	1.0, 1.4	1.1	0.9, 1.3	1672	1.1	0.9, 1.3
Temporal summation: slope of line for first three ratings															
46°C	β coefficient	1495	6.13	0.27	174	6.97	0.77	0.3064	1.1	1.0, 1.3	1.1	0.9, 1.3	1672	1.1	0.9, 1.3
48°C	β coefficient	1485	7.06	0.29	170	7.67	0.85	0.5094	1.1	0.9, 1.3	1.1	0.9, 1.3	1672	1.1	0.9, 1.3
50°C	β coefficient	1430	7.04	0.31	168	8.16	0.86	0.2441	1.1	1.0, 1.3	1.1	0.9, 1.3	1672	1.1	1.0, 1.3

[1] P-values are from analysis of variance model comparing mean values of putative risk factor between cases and controls, with adjustment for study site.

[2] Non-imputed effect estimates are for complete case analysis using numbers of subjects in columns headed n.

[3] Site-adjusted effects were computed in logistic regression models where the putative risk factor as the main explanatory variable and study site as covariate.

[4] Fully-adjusted effects were computed in logistic regression models that additionally include covariates of age group, gender and race/ethnicity.

[5] SOR is standardized odds ratio from a logistic regression model evaluating the linear effect of the risk factor that was standardized by z-score transformation

[6] 95% confidence interval for SOR

[7] Negative z-scores used for standardized odds ratios which therefore represent increase in odds of TMD associated with reduction of 1 sd in the value of the variable

Table 3

Component loadings for principal components analysis model in controls [1]

Pressure pain threshold	Component 1	Component 2	Component 3	Component 4	Component 5
Temporalis	−0.11	−0.05	−0.10	**0.89**	−0.02
Masseter	−0.11	−0.04	−0.07	**0.92**	−0.02
TM joint	−0.11	−0.04	−0.06	**0.90**	−0.03
Trapezius	−0.21	−0.12	−0.10	**0.81**	0.00
Lateral epicondyle	−0.19	−0.05	−0.03	**0.81**	0.01
Mechanical cutaneous pain					
Threshold	−0.17	−0.06	**−0.41**	0.37	−0.07
Single stimulus ratings					
256mN probe	0.26	0.08	**0.71**	−0.11	−0.04
512mN probe512mN probe	0.36	0.10	**0.72**	−0.13	−0.06
Aftersensation ratings					
15 s, 256mN probe	0.01	0.27	**0.85**	0.01	−0.03
30 s, 256mN probe	−0.03	0.28	**0.79**	0.03	−0.02
15 s, 512mN probe	0.11	**0.40**	**0.80**	−0.05	−0.01
30 s, 512mN probe	0.06	**0.41**	**0.76**	−0.05	−0.04
Temporal summation					
256mN probe	0.28	0.09	**0.58**	−0.14	0.00
512mN probe	0.37	0.14	0.31	−0.16	0.10
Heat pain tolerance	**−0.46**	−0.25	−0.12	0.37	−0.08
Single stimulus ratings					
46°C	**0.77**	0.17	0.20	−0.16	−0.36
48°C	**0.85**	0.17	0.18	−0.15	−0.32
50°C	**0.88**	0.14	0.12	−0.13	−0.21
Ratings of 10 stimuli: area under curve					
46°C	**0.86**	0.18	0.19	−0.18	0.11
48°C	**0.91**	0.19	0.16	−0.17	0.19
50°C	**0.83**	0.18	0.14	−0.17	0.27
Thermal aftersensations					
15 s, 46°C	0.21	**0.80**	0.24	−0.05	−0.01

Pressure pain threshold	Component 1	Component 2	Component 3	Component 4	Component 5
30 s, 46°C	0.23	**0.85**	0.25	−0.07	0.03
15 s, 48°C	0.16	**0.84**	0.26	−0.08	0.05
30 s, 48°C	0.14	**0.83**	0.17	−0.04	−0.01
15 s, 50°C	0.17	**0.88**	0.19	−0.09	0.01
30 s, 50°C	0.09	**0.86**	0.24	−0.07	0.04
Temporal summation: highest minus first rating					
46°C	0.10	0.03	−0.02	−0.03	**0.83**
48°C	−0.06	0.02	−0.05	−0.04	**0.87**
50 °C	−0.33	0.02	−0.01	−0.02	**0.72**
Temporal summation: slope of line for first three ratings					
46°C	0.25	0.00	0.04	0.00	**0.70**
48°C	0.10	0.00	−0.03	0.00	**0.80**
50°C	−0.18	0.03	0.01	−0.02	**0.74**
Cumulative variance	0.17	0.32	0.46	0.59	0.72
Cronbach's alpha	0.96	0.95	0.9	0.93	0.87

[1] Data are from 1633 controls who do not have TMD