PREDICTORS OF PERSISTENT TMD PAIN: A 9-YEAR COHORT STUDY

A THESIS SUBMITTED TO THE FACULTY

OF THE UNIVERSITY OF MINNESOTA

BY

FLAVIA PENTEADO KAPOS

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

ERIC LEO SCHIFFMAN

OCTOBER 2015

Acknowledgements

Firstly, I would like to express my gratitude to my advisor Dr. Eric Schiffman for the support of my M.S. research, for his generosity in sharing his data and his great research enterprise. His leadership motivated me to become a better communicator and a stronger researcher. Besides my advisor, I would like to thank the other members of my thesis committee: Dr. John Look and Dr. James Hodges, for their insightful comments and encouragement. My sincere thanks also goes to Dr. Donald Nixdorf and Dr. Estephan Moana, who also provided academic guidance and support, which were fundamental to keep me on track during this journey.

I also thank Patricia Lenton for the help with access to the study archives and with polishing my presentation. Thank you to my fellow residents for sharing this great learning experience with me, and to the University of Minnesota for providing an incredible environment to learn and thrive. At last but not least, I would like to thank my family for believing in me and supporting me unconditionally to pursue my success.

Dedication

This thesis is dedicated to my parents, who devoted their lives to provide me with everything I need: in memory of my mother Glaura and her unstoppable Love, and to my father Roberto, who continuously supports and encourages me to achieve my full potential. I also dedicate this work to my grandmother Aurea, the strongest person I know, in gratitude for her inspiration, and to my brother Guilherme, my eternal best friend. Amo vocês.

Abstract

Aims: To determine, in subjects with painful temporomandibular disorders (TMD), whether baseline SF-12 Physical Component Summary (PCS), SF-12 Mental Component Summary (MCS) and Jaw Functional Limitation Scale (JFLS) predict persistent TMD pain measured by Characteristic Pain Intensity (CPI) scores after 9 years.

Methods: 258 subjects with painful TMD diagnoses and CPI>0 completed baseline SF-12 Health-Related Quality of Life (HRQoL) and JFLS questionnaires. After 9 years, they were reevaluated for painful TMD diagnoses and completed the CPI questionnaire. Univariable and multivariable linear regression adjusted for age and sex examined the relationship between baseline predictors and follow-up CPI.

Results: After 9 years, 186 (72%) had persistent TMD pain. Baseline PCS and JFLS, but not MCS, presented weak linear relationships with follow-up CPI. One SD (9.0) increase in baseline PCS was associated with a 4.9-point decrease in follow-up CPI (SE=1.2, $p<0.001$), or 5.7% of the follow-up CPI score range. One SD (1.4) increase in baseline JFLS was associated with a 5.0-point increase in follow-up CPI scores (SE=1.2, $p<0.001$), or 5.8% of the follow-up CPI score range. In the 3-predictor multivariable model, follow-up CPI change predicted by 1 SD increase in baseline scores was of -4.7 (SE=1.3, $p<0.001$) for PCS, and 3.9 points (SE=1.2, $p=0.002$) for JFLS.

Conclusions: In subjects with TMD pain, baseline PCS and JFLS, but not MCS, were statistically significant predictors of CPI at a 9-year follow-up. However, the magnitude of the effects is small and below the 10-20% minimum change in pain intensity recommended for clinical significance.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

BL	Baseline
CPI	Characteristic Pain Intensity
DC-TMD	Diagnostic Criteria for Temporomandibular Disorders
Dx	Diagnoses
HR	Hazard ratio
HRQoL	Health-related quality of life
IMMPACT	Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials
JFLS	Jaw Functional Limitation Scale
MCS	Mental Component Summary score
MID	Minimal important difference
n	Sample size
NRS	Numerical rating scale
p	p-value
PCS	Physical Component Summary score
r	Pearson correlation coefficient
r^2	Coefficient of determination
RDC-TMD	Research Diagnostic Criteria for Temporomandibular Disorders
SD	Standard deviation
SE	Standard error
SF-12v2	Optum™ SF-12v2 Health Survey
SF-36	Optum™ SF-36 Health Survey
TMD	Temporomandibular disorders
TMJ	Temporomandibular joint
WOMAC	Western Ontario and McMaster Universities Arthritis Index

Introduction

Temporomandibular disorders (TMD) consist of different clinical disorders involving the masticatory muscles, the temporomandibular joints (TMJ) and the associated structures. [1] Painful TMD is the most frequent reason patients seek treatment and is estimated to affect approximately 10% of the adult population.[2] It has been reported that the majority of TMD cases tend to remit or present as recurrent, but approximately 15% of the patients who seek care progress to chronic TMD pain[3] , often defined as pain lasting longer than 3 months. [4] Causal factors for the persistence of TMD pain are not clear, but research has demonstrated that psychosocial factors are significantly higher in TMD patients than pain-free controls[5-8] and are associated with chronic TMD pain.[9-11] Prior studies have suggested that biopsychosocial factors including poor general health can predict onset of chronic pain conditions such as widespread pain[12] and other musculoskeletal disorders[13-17] including orofacial pain and TMD.[18-21]

Health related quality of life (HRQoL) measures have been used to assess the individual's functioning and disease burden for many chronic physical conditions including headaches,[22] arthritis,[23,24] back pain,[15,23,24] and TMD. [6,25,26] The 12-item Short Form Health Survey (SF-12),[27,28] is a commonly used HRQoL questionnaire, derived from the 36-item version, SF-36.[29] The SF-12 and the SF-36 evaluate 8 health domains (physical functioning, role physical, bodily pain, general health, mental health, role emotional, social functioning and vitality), which can be scored as separate sub-scales. The SF-12 and SF-36 also yield 2

summary scores: the Physical Component Summary (PCS) and the Mental Component Summary (MCS). SF-12 and SF-36 scores have been shown to differ significantly between patients and healthy controls for migraine,[30,31] fibromyalgia,[31] and TMD.[25] SF-12 scores have been reported to have a dose-response relationship according to the severity of various chronic health conditions.[32] In a multivariable model, the SF-12 bodily pain and general health sub-scales were among the 11 best predictors for new-onset TMD out of 202 putative risk factors evaluated in a large cohort study. [33] To date, no studies have evaluated the capacity of SF-12 scores to predict pain outcomes in TMD subjects. Simple assessment instruments with predictive value for identifying patients at risk for persistent TMD pain are highly desirable.

According to the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT),[34] there are 2 broad areas of assessment for physical functioning and, more generally, HRQoL: generic global measures, such as the SF-12, and disease-specific measures. The Jaw Functional Limitation Scale (JFLS)[35] assesses TMD-specific physical functioning. The JFLS is a DC-TMD[36] Axis II instrument and its scores have been found to be significantly higher in TMD cases compared to pain-free controls in a large case-control study.[37]

The aim of this study is to determine, in adults with baseline TMD pain, whether baseline PCS, MCS and JFLS scores predict persistent TMD pain at a 9-year follow-up.

Methods

Study sample

This study's baseline and follow-up data are from the multi-center longitudinal observational Validation Project[38] and TMJ IMPACT Project, respectively. Baseline subjects were recruited at the University of Minnesota, University of Washington, and University at Buffalo, between 2003 and 2006. Enrollment was consecutive until two thirds of the previously calculated sample size was achieved, and selective until the closure of the study, to enrich the sample with less frequent TMD diagnoses. Thus the Validation Project subject population was a convenience sample, consisting of clinical and community TMD cases with the full spectrum of TMD signs and symptoms, as well as healthy controls. The Validation Project's methods, including inclusion and exclusion criteria for TMD cases and controls, have been previously reported.[38]

Figure 1 shows the flow diagram of the study sample. 732 subjects enrolled in the Validation Project. 8 subjects dropped out or incomplete assessments, and 724 subjects completed the baseline phase. The study's examiners could not reach a consensus diagnosis for 5 subjects, and 14 subjects were excluded due to reported co-morbid pain conditions (e.g., fibromyalgia). The remaining 705 subjects consisted of 614 TMD cases and 91 controls.[38] Of the 614 TMD cases, 513 subjects had a least one painful TMD diagnosis, including masticatory muscle pain (i.e., myofascial pain with and without limited opening) and/or TMJ pain (i.e., TMJ arthralgia and TMJ osteoarthritis). In the TMJ IMPACT Project,

401 Validation Project subjects who had previously agreed to be contacted were recalled. Of the 401 subjects that returned for follow-up, 273 subjects had at least one baseline painful TMD diagnosis. Of the 273 potential cases at follow-up, 8 were excluded from our analyses due to missing data in the baseline SF-12, 2 were excluded due to having a baseline CPI score of 0, and 4 due to having painful TMD diagnoses but a CPI score of 0. One subject was excluded due to missing data for the follow-up CPI score. Therefore, our total sample included 258 subjects.

For the purpose of this study, TMD pain was defined as at least one painful TMD diagnosis and a Characteristic Pain Intensity (CPI) score greater than 0.[39] Persistent TMD pain was defined as being present at baseline and at follow-up, after 9 years. Inclusion criteria for the present study required complete data for: (1) baseline painful TMD diagnosis, (2) baseline CPI score greater than 0, (3) baseline SF-12 (PCS and MCS) scores, and (4) follow-up CPI scores. Subjects with concurrent painful TMD diagnoses and an inconsistent CPI score of 0 were excluded, since the TMD diagnoses required a report of pain in the last month and a CPI of 0 indicates no pain in the last 6 months.

Clinical assessment

At follow-up, subjects had a complete clinical assessment similar to that used at baseline in the Validation Project.[38] However, one methodological difference between baseline and follow-up was that the painful TMD diagnoses at baseline

were rendered by the consensus of 2 clinical examiners at each site. At follow-up, no consensus diagnosis was rendered as each subject saw just one of the examiners. The painful TMD diagnoses at follow-up were algorithmically derived using the Axis I Diagnostic Criteria for TMD (DC/TMD) protocol.[36] The examiners at baseline and at follow-up were the same individuals at all 3 study sites. Inter-rater reliability of examiners, assessed with kappa, was 0.83 for masticatory muscle pain and 0.85 for TMJ pain at baseline[40] and 0.84 for masticatory muscle pain and 0.76 for TMJ pain at follow-up. (Unpublished data)

Baseline predictors

Baseline questionnaires included the SF-12 version 2[27] and the JFLS-20 (JFLS).[35,41] The SF-12 is a reliable and valid HRQoL questionnaire[28,32,42] used in both the clinical and research setting. The SF-12 MCS and PCS scores range from 0 to 100 points, and a higher score indicates a better quality of life. Cross-validation between the SF-12 and the SF-36 demonstrated high correlation coefficients of 0.95 for the PCS and 0.97 for the MCS. The SF-12 can also be scored in the 8 domain sub-scales; however, their use was not recommended after a validation assessment revealed unsatisfactory correlations with SF-36.[28]

The JFLS is comprised of 20 items each with a 0-10 numerical rating scale of jaw functional limitation, where 0 indicates no jaw limitation and 10 indicates severe jaw limitation in a specified activity. A “n/a” option is also available for each item, and is marked as missing. The JFLS has 3 subscales: mastication, vertical jaw

mobility, and verbal and emotional expression, calculated as the average of all items in that subscale. A global jaw functional limitation score, on a 0-10 scale, is obtained by an average of these three subscales. The JFLS has been validated in TMD patients and the reliability coefficients reported were 0.82 for persons and 0.99 for items.[41] The JFLS was developed through item-analysis of the Seattle Checklist[43] and the Mandibular Functional Impairment Questionnaire[44] to improve psychometric proprieties of validity, reliability and responsiveness for this construct.

Outcome measure

The CPI questions, from the Graded Chronic Pain Scale,[39] were administered at baseline and at follow-up. The primary outcome measure for this study is the follow-up CPI score, which ranges from 0 to 100, and is calculated based on three scales of self-reported pain intensity; a) pain at present time, b) worst pain in the last 6 months and c) average pain in the last 6 months. Each of these scales ranges from 0 to 10, with 0 = “no pain” and 10 = “pain as bad as it could be”. The 3 scores are averaged and multiplied by 10 to obtain the CPI score.

Statistical analyses

Descriptive statistics are presented for baseline and follow-up data. Two-sample t-tests were used for continuous variables and Fisher’s exact tests for categorical variables, to determine if there were differences in mean baseline PCS, MCS and JFLS scores, as well as other baseline characteristics, between groups with and

without follow-up painful TMD diagnoses. Linear regression analysis was used to investigate the relationship between baseline PCS, MCS, JFLS global score and its sub-scales as predictors of follow-up CPI in single variable and multivariable models adjusted for age and sex.

Pearson correlation coefficients and coefficients of determination measured associations of baseline PCS, MCS and JFLS with baseline CPI. Additionally, tertile analyses were used to evaluate the dose-response relationship of baseline PCS, MCS and JFLS global score with follow-up CPI. Tertile analyses were adjusted for age and sex, and corrected by Tukey's method for multiple comparisons. Statistical significance was defined as $p<0.025$ for the two primary analyses (baseline PCS and MCS predicting follow-up CPI), i.e., a Bonferroni correction (0.05/2) was used for the 2 comparisons. For all other tests, considered exploratory, statistical significance was defined as $p<0.05$.

Results

Of the 258 included subjects with TMD pain at baseline, 186 (72%) had painful TMD diagnoses after 9 years. The baseline sample consisted of 88% women, with a mean age of 38 years (standard deviation [SD]=13), ranging from 18 to 67 years. Baseline PCS scores differed significantly between subjects with painful TMD diagnoses at follow-up (PCS=49.8, SD=9.5) and subjects without painful TMD diagnoses at follow-up [(PCS=52.6, SD=7.0), p=0.009]. MCS and all other baseline characteristics did not differ significantly between these two groups, except for the JFLS verbal/emotional sub-scale, where subjects with painful TMD diagnoses at follow-up had significantly higher JFLS verbal/emotional sub-score at baseline (JFLS verbal/emotional=0.9, SD=1.3), compared to subjects without painful TMD diagnoses at follow-up [(JFLS verbal/emotional=0.5, SD=1.1), p=0.045]. Table 1 presents baseline characteristics of subjects with and without painful TMD diagnoses at follow-up are presented.

Table 2 presents follow-up CPI and painful TMD diagnoses. Our primary outcome measure was the follow-up TMD pain intensity per the CPI score, which had a mean of 28.1 and a SD of 19.9 points, with values ranging from 0 to 86.7 points on a 0-100 scale. The baseline CPI scores (mean=50.2, SD=20.0 points) decreased an average 22.0 points at follow-up, which represents a 43.8% decrease in overall TMD pain intensity. At follow-up some subjects reported a CPI score greater than 0 for a 6-month reference time frame, but were classified as not having a painful TMD diagnoses because the clinical diagnoses were

based on pain in the last month. These subjects were not excluded, because this is a clinically possible situation, hence the reason Table 2 shows subjects with CPI greater than 0 in the group without a follow-up painful TMD diagnosis.

Single variable linear regression analyses

Table 3 presents the single variable analyses for baseline PCS, MCS and JFLS scores as predictors of follow-up CPI. Figures 2-4 graphically display these findings. Figure 2 shows the weak negative linear relationship observed between baseline PCS and follow-up CPI. One SD (9.0) increase in baseline PCS scores was associated with a decrease in follow-up CPI scores of 4.9 points (Standard error [SE]=1.2, $p<0.001$). Compared to the range of CPI scores at follow-up (0.0 to 86.7), this was a 5.7% decrease, meaning that a better health-related quality of life was a weak protective factor. Figure 3 shows the relationship between baseline MCS and follow-up CPI, which was not statistically significant. One SD (9.1) increase in baseline MCS scores was associated with a decrease in follow-up CPI scores of 1.4 (SE=1.2, $p=0.258$).

Figure 4 shows the weak positive linear relationship between baseline JFLS global score and follow-up CPI. One SD (1.4) increase in baseline JFLS global score was associated with an increase in follow-up CPI scores of 5.0 points (SE=1.2, $p<0.001$). This represents 5.7% of the range of the follow-up CPI scores. The 3 baseline JFLS sub-scales were statistically significant predictors of

follow-up CPI, but the change they predicted was not greater than predicted by the JFLS global score (see Table 3).

Multivariable linear regression model

The multivariable linear regression model included baseline PCS, MCS and JFLS global scores (Table 4). One SD (9.0) increase in baseline PCS scores was associated with a decrease in follow-up CPI score of 4.7 points (SE=1.3, $p<0.001$), when controlling for the other variables in the model. One SD (1.4) increase in baseline JFLS global score was associated with an increase in follow-up CPI scores of 3.9 points (SE=1.2, $p=0.002$) when controlling for the other variables in the model.

Baseline cross-sectional analyses

In a cross-sectional analysis of the baseline data (Table 5), baseline PCS and baseline had a weak negative correlation CPI ($r=-0.23$, $p<0.001$); baseline PCS explained 5.4% of the variance in baseline CPI ($r^2=0.054$). Baseline MCS also had a very weak negative correlation with baseline CPI ($r=-0.13$, $p=0.043$), explaining only 1.6% of the variance in baseline CPI ($r^2=0.016$). Baseline JFLS global score had a strong positive correlation with baseline CPI ($r=0.53$, $p<0.001$), explaining 27.8% of the variance in baseline TMD pain intensity per CPI.

Tertile analyses

The tertiles of PCS and JFLS global score differed significantly in follow-up CPI scores (Table 6). Figure 5 shows the dose-response relationship between baseline PCS tertiles and follow-up CPI. Lower (worse) baseline PCS scores were associated with higher (worse) CPI scores at follow-up. The 1st tertile, representing individuals with lowest (worst) PCS scores, had a mean follow-up CPI score of 30.5 points (SE=2.6), which was significantly higher than the 2nd (CPI=22.5, SE=2.5, p=0.023) and 3rd (CPI=18.7, SE=2.4, p<0.001) tertiles. The 1st and 3rd tertiles differed by 11.7 points in mean follow-up CPI score This represents 13.5% of the follow-up CPI score range. The 7.9-point difference between 1st and 2nd tertiles is 9.1% of the follow-up CPI score range. Figure 6 shows the relationship between baseline MCS tertiles and follow-up CPI; MCS tertiles did not differ significantly in follow-up CPI (p>0.05).

Figure 7 shows the dose-response relationship between baseline JFLS global score tertiles and follow-up CPI. Higher (worse) baseline JFLS global scores were associated with higher (worse) CPI scores at follow-up. The 1st tertile, representing individuals with the best jaw function, had a mean follow-up CPI score of 17.2 points (SE=2.4), which was significantly lower than the 2nd (CPI=25.1, SE=2.4, p=0.019) and 3rd (CPI=31.1, SE=2.6, p<0.001) tertiles. The 1st and 3rd tertiles differed by 13.9 points in mean follow-up CPI score. This represents 16.0% of the follow-up CPI score range. The 7.9-point difference between 1st and 2nd tertiles is 9.1% of the follow-up CPI score range (Table 6).

Discussion

The present study found a statistically significant association between baseline PCS, MCS and JFLS scores, and follow-up CPI. However, the effects were too small to be clinically significant. The correlations between the variables in our study are consistent with previous reports.[18,20,33,45] Nevertheless, this is the first study to assess the clinical significance of these findings. Our results also cohere with the current concept of a complex multifactorial etiologic pattern for TMD,[45] where single variables are unlikely to have robust prognostic value.

Clinical Significance

Assessing and interpreting the clinical significance of research findings is extremely important to bridge the gap between research and clinical practice. IMMPACT has discussed and reviewed different strategies to determine clinically significant changes in outcomes for chronic pain. Recommendations based on a literature review suggest that a decrease of 10-20% in self-reported pain intensity reflects minimally important improvement, 30% or greater reflects moderately important improvement, and 50% reflects substantial improvement.[46] In addition, data-driven estimates found that a decrease of 2 points or approximately 30% on a 0-10 numerical rating scale is considered clinically important to patients with osteoarthritis, fibromyalgia, chronic low-back pain, diabetic neuropathy and post-herpetic neuralgia.[47] For acute pain studies, 50% improvement in pain intensity has been suggested as the cutoff to evaluate clinical efficacy of analgesics. [48]

SF-12 literature review

No previous studies have used baseline SF-12 to predict follow-up pain intensity in TMD subjects. A study evaluating multiple putative risk factors for first-onset TMD[18] found statistically significant adjusted hazard ratios (HR) for 1 SD increase in baseline PCS (HR [95%CI]=0.85 [0.74, 0.95], p=0.008) and MCS (HR [95%CI]=0.74 [0.66, 0.82], p<0.001). This means that an increase of 1SD in baseline PCS scores, decrease the probability of an individual developing first-onset TMD during the 2.8-year follow-up by 15%. The individual probability is decrease by 26% with an increase of 1 SD in baseline MCS scores. The clinical significance of these findings was not discussed, but the authors did not include SF-12 measures in multivariable models of general-health status as predictor of first-onset TMD because it was considered a consequence of other health-conditions, rather than an independent predictor. [18] In the present study, a 1 SD increase in baseline PCS score predicted a reduction in follow-up CPI scores of 4.9 points (5.7%), which is below suggested standards for clinical significance, despite its statistical significance. The change predicted by MCS was less than one-third of that observed for PCS and was not statistically or clinically significant.

Prior studies have reported a minimal important difference (MID) in the summary scores of the SF-12 and SF-36 questionnaires as low as 2-5 points.[49-51] Researchers proposing indexed or distribution-based MIDs have found different results. Recommendations include defining MID as 0.5 SD or 1.96 SE, based on

observed patterns in the literature, the statistical concept of sample variability, and research in psychology that measures the human ability to discriminate changes.[52-54] Our study found a difference in mean baseline PCS that was statistically significant between groups with and without follow-up painful TMD diagnoses. A difference of 3.1 points was found in the normalized 0-100 scale of the SF-12, which represents 6.1% of the score range (15.7 to 65.4 points), or less than 0.5 SD, and is not considered clinically significant (Table 1).

Two cross-sectional studies of subjects with TMD found differences in 3 of the 8 SF-36 sub-scores, when compared to controls. In the first, general health, vitality, and social functioning were statistically significant.[24] In the second, physical functioning, bodily pain and social functioning were statistically significant.[6] PCS and MCS scores were not reported. The scoring of the sub-scales, as in the summary scores, is based on a 0-100-point scale, where higher scores reflect better HRQoL. The magnitude of the differences between groups ranged from 9.1 to 20.8 points. The differences found in these other studies are greater than in the present study, which may be due to comparing TMD subjects with subjects from the general population[25] or age and sex matched pain-free controls,[6] while our entire sample had TMD pain at baseline.

JFLS literature review

IMMPACT recommends the use of disease-specific measures of physical functioning when available.[33] Disease-specific instruments can assess aspects of

physical functioning and pain-imposed limitations associated with the studied disorder, disease, or condition that would not be addressed by generic measures. Examples of such instruments are the Western Ontario and McMaster Universities Arthritis Index (WOMAC)[55] and the Roland and Morris Back Pain Disability Scale.[56] In this sense, the JFLS is an appropriate instrument to assess TMD-specific physical functioning.

The JFLS scores were added *a posteriori* in this study and analyzed as a potential predictor of persistent pain, to include disease-specific data in our analysis. The JFLS has been found to be significantly associated with TMD when TMD cases were compared to pain-free controls.[37] That study found that for every 1 SD increase in JFLS sub-scores, the odds of being in the TMD group of the case-control study increased by 3.0 times for mastication and vertical jaw mobility sub-scales (OR [95%CI] = 3.0 [2.6, 3.5]), by 1.6 times for the verbal/ emotional expression sub-scale (OR [95%CI] = 1.6 (1.4, 1.8) and by 2.9 times for the JFLS global score (OR [95%CI] = 2.9 (2.5, 3,4). Mean JFLS scores for cases in this study[37] were very similar to the present study. The authors reported mean scores of 2.22, 2.22, 0.72 and 1.74 points for mastication, opening, verbal/emotional expression, and global score, respectively, compared to mean scores of 2.33, 2.50, 0.79 and 1.84 points, respectively, in the present study. JFLS scores were not significant predictors of new-onset TMD in a follow-up cohort study done by the same group.[57]

This is the first study to report findings on the JFLS as a predictor of persistent TMD pain. Baseline JFLS global score and its sub-scales were statistically significant predictors of follow-up CPI in a linear regression analysis adjusted for age and sex. Nevertheless, the magnitude of the association was effectively the same as for baseline PCS, for which a 1 SD change was associated with a 5.7% change in follow-up CPI scores for the JFLS global score. As argued above, these likely do not represent large enough effects to be clinically significant.

Multivariable analysis

When baseline PCS, MCS and JFLS global scores were combined in a multivariable model, PCS and JFLS remained statistically significant. Especially for the baseline PCS score, the effect size was largely maintained, changing from -4.9 to -4.7 when controlling for the other variables in the model. The stability of the effect despite adjusting for potential confounders can be interpreted as a sign that the observed effect is in fact present. The magnitude and the direction of change in effect after partial adjustment for confounding can be used to make inferences about the actual effect.[58] Nonetheless, it reinforces the lack of clinical significance, because the true effect would be highly unlikely to move in the opposite direction, towards a greater magnitude.[58]

Sample selection

The original Validation Project was not designed to address the aims of the present study, but it followed specific guidelines: the statement for reporting

studies of diagnostic accuracy (STARD).[59] These guidelines recommend that, when developing diagnostic criteria, the first step is to use a sample of individuals with the target condition and free of significant co-morbidities. According to the trend observed in the tertile analyses in the present study, subjects with worse baseline PCS and JFLS scores presented a greater change in follow-up CPI scores. If that trend extended beyond the range of our data, it is possible that the effect sizes would reach clinical significance in subjects with higher baseline impairment in HRQoL and jaw function. While it seems plausible that, in subjects with non-TMD pain and other medical co-morbidities, baseline PCS and JFLS scores could potentially be worse, the predictive value of baseline PCS and JFLS for follow-up CPI in such population is unknown.

9- year follow-up and scope of evaluation

The long interval between baseline and follow-up data may account for difficulty in predicting the outcome of persistent pain. The status of subjects during this follow-up period is unknown, and treatments received between evaluations were not considered in the analyses. It has been suggested that a 1-month recall period allows for a better self-reported pain intensity rating,[60] but the 6-month period is also considered acceptable.[42,61] In a study of migraineurs, self-report of pain intensity (0-10 scale) over the 3 previous months had a high correlation (r=0.74) with a daily pain diary, which was considered the reference standard.[62]

In our case, the 6-month reference was used since this was the period used in the original questionnaire[38] and in the Validation Project.[38] Also, considering the fluctuating nature of TMD pain, a 6-month reference period is thought to provide a better coverage of the time between visits, especially if the interval between follow-ups is over a long period like the present study. The long interval between the two measurement times introduces a level of uncertainty; however, the baseline cross-sectional analysis shows that stronger associations may never have been present (Table 5). The strength of the associations between putative baseline predictors and baseline CPI were not dramatically different than in the longitudinal analyses, which supports the results of this study.

The TMJ IMPACT Project, which provided the follow-up data for this study, recalled only 401 subjects who had previously agreed to be contacted for reevaluation, out of the 614 cases and 91 controls from the Validation Project. Thus, selective follow-up may have introduced bias into the analysis. Although 15 of 273 subjects were excluded due to inconsistent or missing data, these represent only 5% of the potential sample and are unlikely to introduced major bias into the results. Finally, the present study did not account for a series of other potential confounding factors that could be used in the multivariable analysis for persistent TMD pain, such as TMD pain chronicity, oral behaviors, somatization, depression and anxiety.[36]

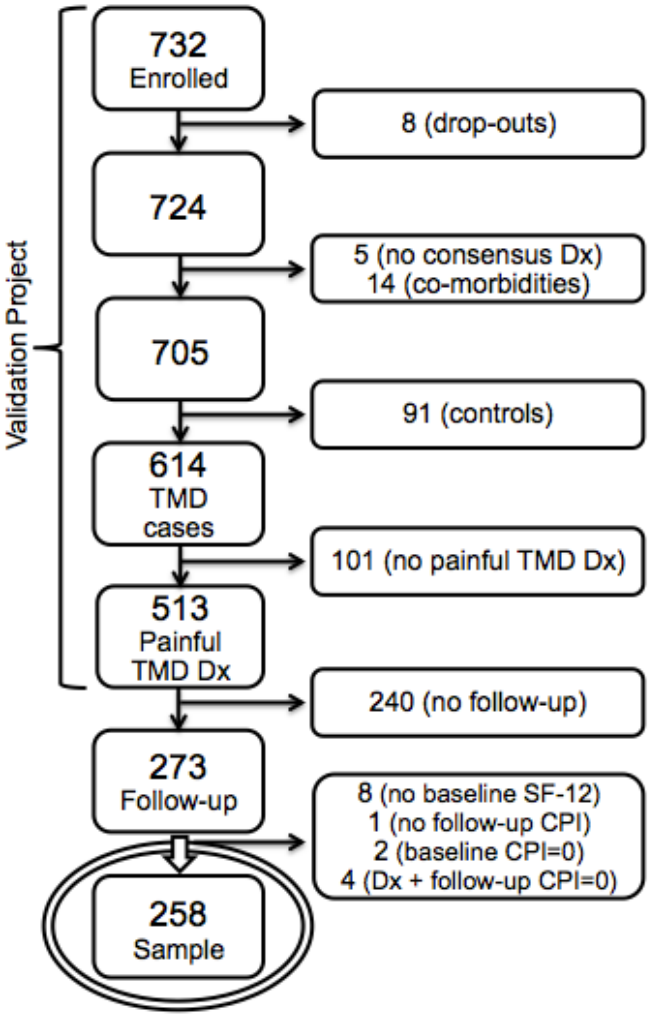
When evaluating patient reported outcomes (PROs) of jaw pain, jaw function and quality of life, these outcomes should be self-reported, without interpretation of the patient's response by a clinician or anyone else.[63] PROs, and specifically HRQoL, have become important constructs in patient-centered research and clinical practice. Some authors argue that "[t]he ultimate goal of health care is to restore or preserve functioning and well-being related to health, that is health-related quality of life."[64] Nevertheless, assessment of PROs remains inherently subject to the limitations of the instruments that measure them. If the instrument does not measure the desired construct adequately, then non-significant study results could arise from its shortcomings. Validity assessment of questionnaires typically involves construct and not criterion validity.[65] Construct validity is based on how consistently a particular measure relates to theoretical hypotheses about the concepts being measured.[66] For instance, subjects with more severe health conditions, presumed to have higher HRQoL impairment, should consistently obtain worse HRQoL scores than healthy subjects or subjects with milder health conditions. The SF-12 construct validity has been assessed by comparing scores of subjects in different groups. SF-12 scores were able to discriminate between these groups, known to differ in presence and seriousness of physical and mental health conditions.[28,32] Concurrent validity for the SF-12 was also established versus the SF-36 for PCS and MCS,[28,32] and versus the EuroQol EQ-5D, another HRQoL questionnaire.[58]

Conclusions

This is the first study to examine both the statistical and clinical significance of the longitudinal association of HRQoL and jaw functional limitation with persistent TMD pain intensity. In subjects with TMD pain, baseline PCS and JFLS scores were statistically significant predictors of CPI at a 9-year follow-up. However, the magnitude of the effects is small and not clinically significant. Future studies should routinely report the clinical significance of their findings to bridge the current gap between research findings and their application in the clinical setting.

Illustrations

Figure 1: Flow Diagram of Study Subjects



CPI: Characteristic Pain Intensity, Dx: Diagnosis.

Table 1: Baseline Characteristics of Subjects With and Without Follow-up Painful TMD Diagnoses

Variable	Category	Total	Follow-up Painful TMD Dx	Follow-up No Painful TMD Dx	p-value
		(n=258)	(n=186)	(n=72)	
Sex	Male n(%)	31 (12.0)	18 (9.7)	13 (18.1)	0.086
	Female n(%)	227 (88.0)	168 (90.3)	59 (81.9)	
Age (years)	Mean (SD)	37.8 (13.0)	38.4 (13.2)	36.2 (12.7)	0.231
	(Min, Max)	(18.0, 67.0)	(18.0, 67.0)	(18.0, 65.0)	
Baseline CPI 0-100	Total n	258	186	72	0.322
	Mean (SD)	50.2 (20.0)	51.0 (19.5)	48.1 (21.4)	
	(Min, Max)	(6.7, 100.0)	(10.0, 93.3)	(6.7, 100.0)	
Baseline SF-12 PCS 0-100	Total n	258	186	72	0.009
	Mean (SD)	50.6 (8.9)	49.8 (9.5)	52.6 (7.0)	
	(Min, Max)	(15.7, 65.4)	(24.1, 65.4)	(15.7, 64.5)	
Baseline SF-12 MCS 0-100	Total n	258	186	72	0.166
	Mean (SD)	49.6 (9.1)	49.1 (9.21)	50.8 (8.7)	
	(Min, Max)	(22.8, 67.3)	(22.8, 65.2)	(23.3, 67.3)	
Baseline Painful TMD Diagnosis n(%)	Total n	258	186	72	0.353
	Articular Pain	35 (13.6)	26 (14.0)	9 (12.5)	
	Muscular Pain	3 (1.2)	1 (0.5)	2 (2.8)	
	Both	220 (85.3)	159 (85.5)	61 (84.7)	
Baseline JFLS Global Score 0-10	Total n	249	181	68	0.165
	Mean (SD)	1.8 (1.4)	1.9 (1.4)	1.6 (1.5)	
	(Min, Max)	(0.0, 8.0)	(0.0, 6.5)	(0.00, 8.0)	
Baseline JFLS Mastication Sub-scale (0-10)	Total n	257	186	71	0.393
	Mean (SD)	2.3 (1.8)	2.4 (1.7)	2.2 (1.9)	
	(Min, Max)	(0.0, 8.3)	(0.0, 8.3)	(0.0, 7.8)	
Baseline JFLS Opening Sub-scale (0-10)	Total n	255	184	71	0.779
	Mean (SD)	2.5 (2.0)	2.5 (1.9)	2.6 (2.3)	
	(Min, Max)	(0.0, 9.0)	(0.0, 7.3)	(0.0, 9.0)	
Baseline JFLS Verbal/Emotional Sub-scale (0-10)	Total n	250	182	68	0.045
	Mean (SD)	0.8 (1.2)	0.9 (1.3)	0.5 (1.1)	
	(Min, Max)	(0.0, 7.3)	(0.0, 7.3)	(0.0, 7.1)	

CPI: Characteristic Pain Intensity, PCS: Physical Component Summary, MCS: Mental Component Summary, JFLS: Jaw Functional Limitation Scale, Dx: Diagnosis. SD: Standard Deviation.

Table 2: Follow-up Characteristic Pain Intensity and Painful TMD Diagnoses

Variable	Category	Total	Follow-up Painful TMD Dx	Follow-up No Painful TMD Dx
		(n=258)	(n=186)	(n=72)
Follow-up CPI 0-100	Total n	258	186	72
	Mean (SD)	28.1 (19.9)	34.5 (18.0)	11.6 (14.2)
	(Min, Max)	(0.0, 86.7)	(6.7, 86.7)	(0.0, 66.7)
Follow-up Painful TMD Diagnoses n(%)	Total n	258	186	72
	None	72 (27.9)	0	72 (100.0)
	TMJ Pain Only	35 (13.6)	35(18.8)	0
	Muscle Pain Only	5 (1.9)	5 (2.7)	0
	Both	146 (56.6)	146 (78.5)	0

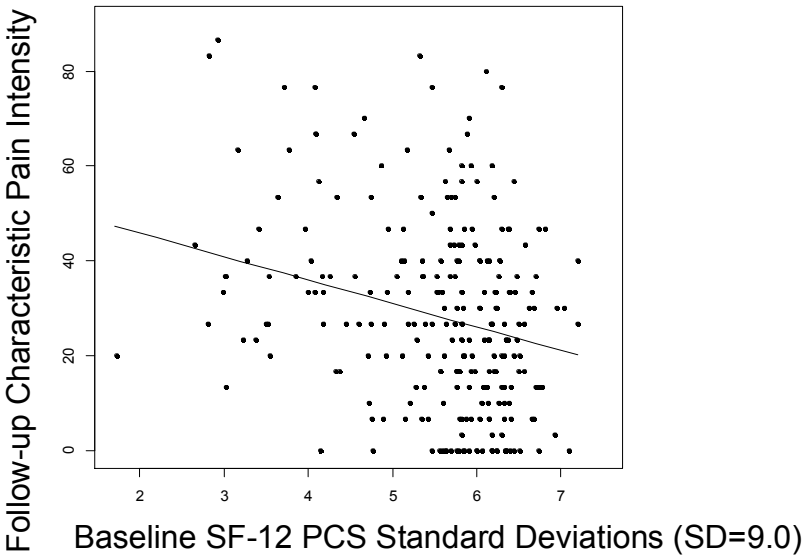
CPI: Characteristic Pain Intensity, Dx: Diagnosis.

Table 3: Change in Follow-up Characteristic Pain Intensity Predicted by 1 SD Increase in Baseline Predictors

Change in Follow-up CPI (Single Variable Analyses)			
Predictor	**Comparison**	**Estimate (SE)**	**p-value**
Baseline SF-12 PCS	1 SD (9.0)	-4.9(1.2)	<0.001
Baseline SF-12 MCS	1 SD (9.1)	-1.4(1.2)	0.258
Baseline JFLS Global Score	1 SD (1.4)	5.0 (1.2)	<0.001
Baseline JFLS Mastication Sub-scale	1 SD (1.8)	4.3(1.2)	<0.001
Baseline JFLS Opening Sub-scale	1 SD (2.0)	2.6(1.2)	0.035
Baseline JFLS Verbal/Emotional Sub-scale	1 SD (1.2)	3.0(0.8)	<0.001

CPI: Characteristic Pain Intensity, PCS: Physical Component Summary, MCS: Mental Component Summary, JFLS: Jaw Functional Limitation Scale, SD: Standard Deviation, SE: Standard Error.

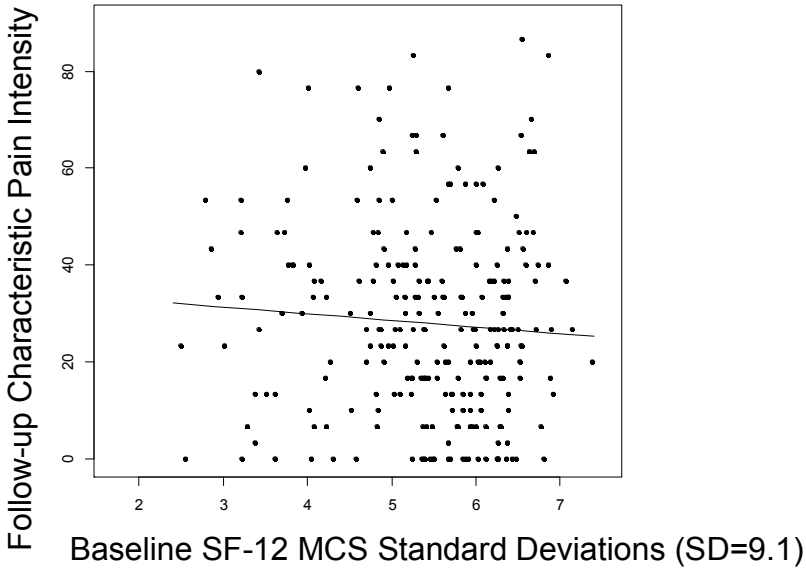
Figure 2: Linear Relationship Between Baseline SF-12 Physical Component Summary and Follow-up Characteristic Pain Intensity



1 SD (9.0) increase in baseline PCS was associated with a 4.9-point decrease in follow-up CPI scores [(SE=1.2), p<0.001].

PCS: Physical Component Summary, CPI: Characteristic Pain Intensity, SD: Standard Deviation, SE: Standard Error.

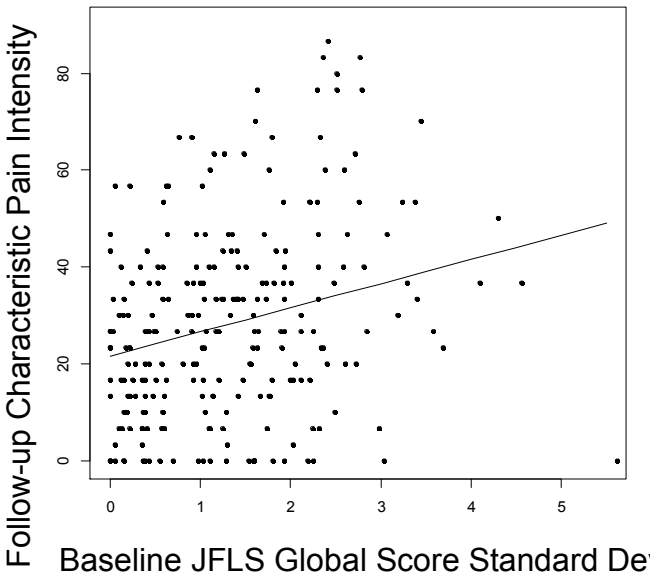
Figure 3: Linear Relationship Between Baseline SF-12 Mental Component Summary and Follow-up Characteristic Pain Intensity



There was not a significant linear relationship between baseline MCS and follow-up CPI (p=0.258).

MCS: Mental Component Summary, CPI: Characteristic Pain Intensity, SD: Standard Deviation.

Figure 4: Linear Relationship Between Baseline Jaw Functional Limitation Scale and Follow-up Characteristic Pain Intensity



1 SD (1.4) increase in baseline JFLS scores was associated with a 5.0-point increase in follow-up CPI scores [(SE=1.2), p<0.001].

JFLS: Jaw Functional Limitation Scale, CPI: Characteristic Pain Intensity, SD: Standard Deviation, SE: Standard Error.

Table 4: Multivariable Model for Predicting Follow-up Characteristic Pain Intensity

Change in Follow-up CPI (Multivariable Analysis)			
Predictor	**Comparison**	**Estimate (SE)**	**p-value**
Baseline SF-12 PCS	1 SD (9.0)	-4.7 (1.3)	<0.001
Baseline SF-12 MCS	1 SD (9.1)	-2.2 (1.3)	0.077
Baseline JFLS Global Score	1 SD (1.4)	3.9 (1.2)	0.002

Estimates and p-values reflect the effect of 1 SD increase in baseline predictors when controlling for all other variables in the model.

CPI: Characteristic Pain Intensity, PCS: Physical Component Summary, MCS: Mental Component Summary, JFLS: Jaw Functional Limitation Scale, SD: Standard Deviation, SE: Standard Error.

Table 5: Association between Baseline Physical Component Summary, Mental Component Summary and Jaw Functional Limitation Scale Global Score, and Baseline Characteristic Pain Intensity

Pearson Correlation Coefficient (r) with Baseline CPI				
Variable	**n**	**r**	**p-value**	**r^2**
Baseline SF-12 PCS	258	-0.23	<0.001	0.054
Baseline SF-12 MCS	258	-0.13	0.043	0.016
Baseline JFLS Global Score	249	0.53	<0.001	0.278

PCS: Physical Component Summary, MCS: Mental Component Summary, JFLS: Jaw Functional Limitation Scale, CPI: Characteristic Pain Intensity, r: Pearson Correlation Coefficient, r^2: Coefficient of Determination.

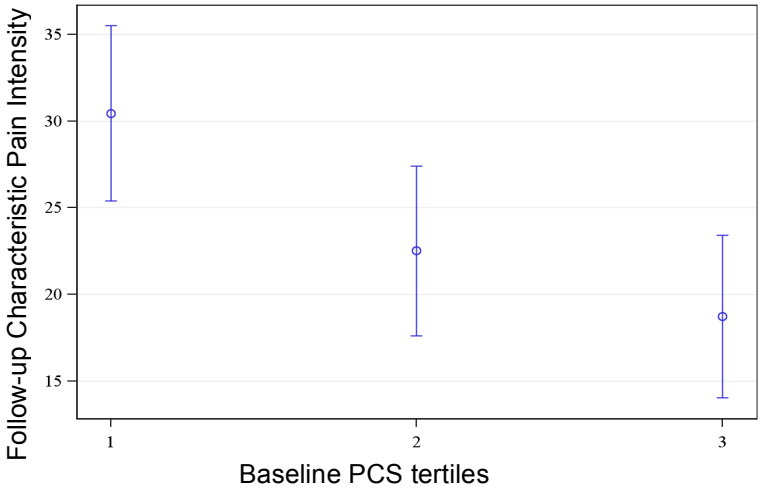
Table 6: Mean Follow-up Characteristic Pain Intensity by Baseline Physical Component Summary, Mental Component Summary and Jaw Functional Limitation Scale Global Score Tertiles

Mean Follow-up CPI (SE)				p-value		
Predictor	**1st tertile**	**2nd tertile**	**3rd tertile**	**1st vs. 2nd**	**1st vs. 3rd**	**2nd vs. 3rd**
SF-12 PCS Tertiles	30.5 (2.6)	22.5 (2.5)	18.7 (2.4)	0.023	<0.001	0.379
SF-12 MCS Tertiles	27.2 (2.5)	21.2 (2.5)	22.4 (2.5)	0.103	0.241	0.904
JFLS Global Score Tertiles	17.2 (2.4)	25.1 (2.4)	31.1 (2.6)	0.019	<0.001	0.107

Lower tertiles represent worse quality of life for PCS/MCS and lower jaw functional limitation (better jaw function) for JFLS.

CPI: Characteristic Pain Intensity, PCS: Physical Component Summary, MCS: Mental Component Summary, JFLS: Jaw Functional Limitation Scale, SE: Standard Error.

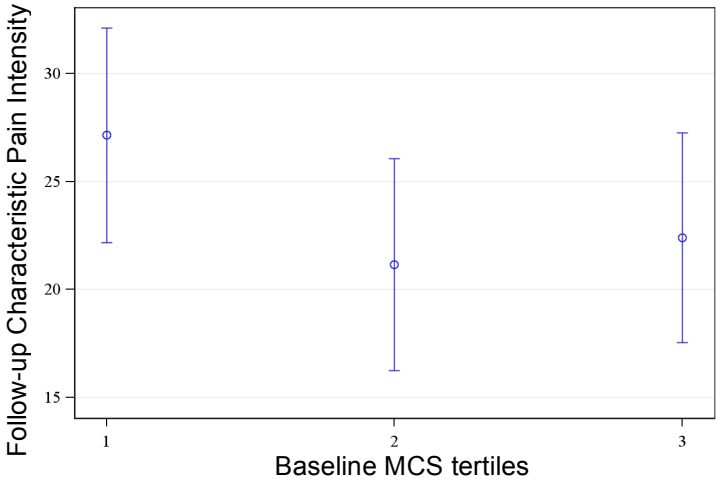
Figure 5: Mean Follow-up Characteristic Pain Intensity (95% CI) by Baseline SF-12 Physical Component Summary Tertiles



Differences between 1st and 3rd (p<0.001), as well as 1st and 2nd (p=0.023) tertiles were statistically significant.

CPI: Characteristic Pain Intensity, PCS: Physical Component Summary, CI: Confidence Interval.

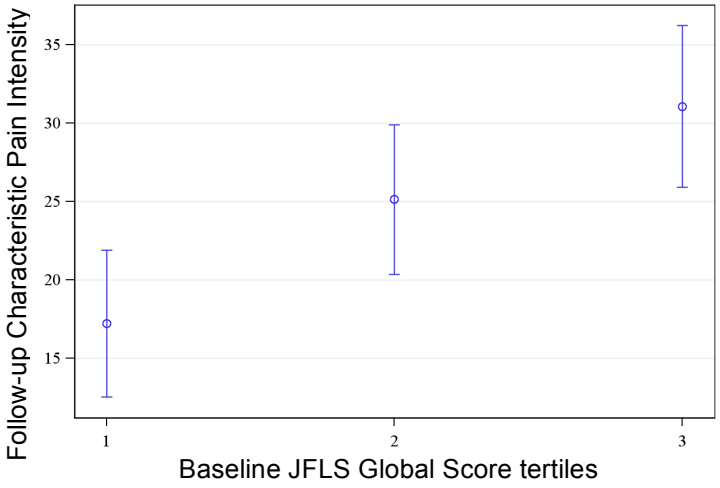
Figure 6: Mean Follow-up Characteristic Pain Intensity (95% CI) by Baseline SF-12 Mental Component Summary Tertiles



There were no statistically significant differences between tertiles. (p>0.05)

CPI: Characteristic Pain Intensity, MCS: Mental Component Summary, CI: Confidence Interval.

Figure 7: Mean Follow-up Characteristic Pain Intensity (95% CI) by Baseline Jaw Functional Limitation Scale Global Score Tertiles



Differences between 1st and 3rd (p<0.001), as well as 1st and 2nd (p=0.019) tertiles were statistically significant.

CPI: Characteristic Pain Intensity, JFLS: Jaw Functional Limitation Scale, CI: Confidence Interval.

Bibliography

1. Okeson JP, de Kanter RJ. Temporomandibular disorders in the medical practice. *J Fam Pract*. 1996;43(4):347-356.

2. LeResche L. Epidemiology of temporomandibular disorders: Implications for the investigation of etiologic factors. *Crit Rev Oral Biol Med*. 1997;8(3):291-305.

3. NIDCR - National Institute of Dental and Craniofacial Research. Find data by topic: Facial pain. http://www.nidcr.nih.gov/DataStatistics/FindDataByTopic/FacialPain/. Updated 20142014.

4. NIH Medline Plus. Chronic pain: Symptoms, diagnosis, & treatment. https://www.nlm.nih.gov/medlineplus/magazine/issues/spring11/articles/spring11pg5-6.html. Updated Spring, 2011. Accessed 10/21, 2015.

5. Quartana PJ, Buenaver LF, Edwards RR, Klick B, Haythornthwaite JA, Smith MT. Pain catastrophizing and salivary cortisol responses to laboratory pain testing in temporomandibular disorder and healthy participants. *J Pain*. 2010;11(2):186-194. doi: 10.1016/j.jpain.2009.07.008 [doi].

6. Johnson J, Johansson M, Ryden A, Houltz E, Finizia C. The impact of trismus on health-related quality of life and mental health. *Head Neck*. 2014. doi: 10.1002/hed.23816 [doi].

7. Dahlstrom L, Carlsson GE. Temporomandibular disorders and oral health-related quality of life. A systematic review. *Acta Odontol Scand*. 2010;68(2):80-85. doi: 10.3109/00016350903431118 [doi].

8. De Leeuw R, Bertoli E, Schmidt JE, Carlson CR. Prevalence of traumatic stressors in patients with temporomandibular disorders. *J Oral Maxillofac Surg*. 2005;63(1):42-50. doi: S027823910401273X [pii].

9. Epker J, Gatchel RJ. Coping profile differences in the biopsychosocial functioning of patients with temporomandibular disorder. *Psychosom Med*. 2000;62(1):69-75.

10. Zakrzewska JM. Multi-dimensionality of chronic pain of the oral cavity and face. *J Headache Pain*. 2013;14(1):37-2377-14-37. doi: 10.1186/1129-2377-14-37 [doi].

11. Velly AM, Look JO, Carlson C, et al. The effect of catastrophizing and depression on chronic pain--a prospective cohort study of temporomandibular muscle and joint pain disorders. *Pain*. 2011;152(10):2377-2383. doi: 10.1016/j.pain.2011.07.004 [doi].

12. McBeth J, Macfarlane GJ, Benjamin S, Silman AJ. Features of somatization predict the onset of chronic widespread pain: Results of a large population-based study. *Arthritis Rheum*. 2001;44(4):940-946. doi: 10.1002/1529-0131(200104)44:4<940::AID-ANR151>3.0.CO;2-S [doi].

13. Lacey RJ, Lewis M, Sim J. Piecework, musculoskeletal pain and the impact of workplace psychosocial factors. *Occup Med (Lond)*. 2007;57(6):430-437. doi: kqm048 [pii].

14. Linton SJ. A review of psychological risk factors in back and neck pain. *Spine (Phila Pa 1976)*. 2000;25(9):1148-1156.

15. Melloh M, Elfering A, Egli Presland C, et al. Identification of prognostic factors for chronicity in patients with low back pain: A review of screening instruments. *Int Orthop*. 2009;33(2):301-313. doi: 10.1007/s00264-008-0707-8 [doi].

16. Walton DM, Carroll LJ, Kasch H, et al. An overview of systematic reviews on prognostic factors in neck pain: Results from the international collaboration on neck pain (ICON) project. *Open Orthop J*. 2013;7:494-505. doi: 10.2174/1874325001307010494 [doi].

17. Zakrzewska JM. Differential diagnosis of facial pain and guidelines for management. *Br J Anaesth*. 2013;111(1):95-104. doi: 10.1093/bja/aet125 [doi].

18. Sanders AE, Slade GD, Bair E, et al. General health status and incidence of first-onset temporomandibular disorder: The OPPERA prospective cohort study. *J Pain*. 2013;14(12 Suppl):T51-62. doi: 10.1016/j.jpain.2013.06.001 [doi].

19. Fillingim RB, Ohrbach R, Greenspan JD, et al. Psychological factors associated with development of TMD: The OPPERA prospective cohort study. *J Pain*. 2013;14(12 Suppl):T75-90. doi: 10.1016/j.jpain.2013.06.009 [doi].

20. Fillingim RB, Ohrbach R, Greenspan JD, et al. Potential psychosocial risk factors for chronic TMD: Descriptive data and empirically identified domains from the OPPERA case-control study. *J Pain*. 2011;12(11 Suppl):T46-60. doi: 10.1016/j.jpain.2011.08.007 [doi].

21. Aggarwal VR, Macfarlane GJ, Farragher TM, McBeth J. Risk factors for onset of chronic oro-facial pain--results of the north cheshire oro-facial pain prospective population study. *Pain*. 2010;149(2):354-359. doi: 10.1016/j.pain.2010.02.040 [doi].

22. Lipton RB, Liberman JN, Kolodner KB, Bigal ME, Dowson A, Stewart WF. Migraine headache disability and health-related quality-of-life: A population-based case-control study from england. *Cephalalgia*. 2003;23(6):441-450. doi: 546 [pii].

23. Picavet HS, Hoeymans N. Health related quality of life in multiple musculoskeletal diseases: SF-36 and EQ-5D in the DMC3 study. *Ann Rheum Dis*. 2004;63(6):723-729. doi: 10.1136/ard.2003.010769 [doi].

24. Salaffi F, De Angelis R, Stancati A, Grassi W, MArche Pain, Prevalence INvestigation Group (MAPPING) study. Health-related quality of life in multiple musculoskeletal conditions: A cross-sectional population based epidemiological study. II. the MAPPING study. *Clin Exp Rheumatol*. 2005;23(6):829-839. doi: 1720 [pii].

25. Tjakkes GH, Reinders JJ, Tenvergert EM, Stegenga B. TMD pain: The effect on health related quality of life and the influence of pain duration. *Health Qual Life Outcomes*. 2010;8:46-7525-8-46. doi: 10.1186/1477-7525-8-46 [doi].

26. Barros Vde M, Seraidarian PI, Cortes MI, de Paula LV. The impact of orofacial pain on the quality of life of patients with temporomandibular disorder. *J Orofac Pain*. 2009;23(1):28-37.

27. Optum™. SF-12® Health survey. https://www.optum.com/optum-outcomes/what-we-do/health-surveys/sf-12v2-health-survey.html.

28. Ware J,Jr, Kosinski M, Keller SD. A 12-item short-form health survey: Construction of scales and preliminary tests of reliability and validity. *Med Care*. 1996;34(3):220-233.

29. Ware JE,Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. conceptual framework and item selection. *Med Care*. 1992;30(6):473-483.

30. Lipton RB, Hamelsky SW, Kolodner KB, Steiner TJ, Stewart WF. Migraine, quality of life, and depression: A population-based case-control study. *Neurology*. 2000;55(5):629-635.

31. Mas AJ, Carmona L, Valverde M, Ribas B, EPISER Study Group. Prevalence and impact of fibromyalgia on function and quality of life in individuals from the general population: Results from a nationwide study in spain. *Clin Exp Rheumatol*. 2008;26(4):519-526. doi: 2383 [pii].

32. Cheak-Zamora NC, Wyrwich KW, McBride TD. Reliability and validity of the SF-12v2 in the medical expenditure panel survey. *Qual Life Res*. 2009;18(6):727-735. doi: 10.1007/s11136-009-9483-1 [doi].

33. Bair E, Ohrbach R, Fillingim RB, et al. Multivariable modeling of phenotypic risk factors for first-onset TMD: The OPPERA prospective cohort study. *J Pain*. 2013;14(12 Suppl):T102-15. doi: 10.1016/j.jpain.2013.09.003 [doi].

34. Dworkin RH, Turk DC, Farrar JT, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain*. 2005;113(1-2):9-19. doi: S0304-3959(04)00440-3 [pii].

35. Ohrbach R, Larsson P, List T. The jaw functional limitation scale: Development, reliability, and validity of 8-item and 20-item versions. *J Orofac Pain*. 2008;22(3):219-230.

36. Schiffman E, Ohrbach R, Truelove E, et al. Diagnostic criteria for temporomandibular disorders (DC/TMD) for clinical and research applications: Recommendations of the international RDC/TMD consortium network* and orofacial pain special interest groupdagger. *J Oral Facial Pain Headache*. 2014;28(1):6-27. doi: 10.11607/jop.1151 [doi].

37. Ohrbach R, Fillingim RB, Mulkey F, et al. Clinical findings and pain symptoms as potential risk factors for chronic TMD: Descriptive data and empirically identified domains from the

OPPERA case-control study. *J Pain*. 2011;12(11 Suppl):T27-45. doi: 10.1016/j.jpain.2011.09.001 [doi].

38. Schiffman EL, Truelove EL, Ohrbach R, et al. The research diagnostic criteria for temporomandibular disorders. I: Overview and methodology for assessment of validity. *J Orofac Pain*. 2010;24(1):7-24.

39. Von Korff M, Ormel J, Keefe FJ, Dworkin SF. Grading the severity of chronic pain. *Pain*. 1992;50(2):133-149. doi: 0304-3959(92)90154-4 [pii].

40. Schiffman EL, Ohrbach R, Truelove EL, et al. The research diagnostic criteria for temporomandibular disorders. V: Methods used to establish and validate revised axis I diagnostic algorithms. *J Orofac Pain*. 2010;24(1):63-78.

41. Ohrbach R, Granger C, List T, Dworkin S. Preliminary development and validation of the jaw functional limitation scale. *Community Dent Oral Epidemiol*. 2008;36(3):228-236. doi: 10.1111/j.1600-0528.2007.00397.x [doi].

42. Ohrbach R, Turner JA, Sherman JJ, et al. The research diagnostic criteria for temporomandibular disorders. IV: Evaluation of psychometric properties of the axis II measures. *J Orofac Pain*. 2010;24(1):48-62.

43. Dworkin SF, LeResche L. Research diagnostic criteria for temporomandibular disorders: Review, criteria, examinations and specifications, critique. *J Craniomandib Disord*. 1992;6(4):301-355.

44. Stegenga B, de Bont LG, de Leeuw R, Boering G. Assessment of mandibular function impairment associated with temporomandibular joint osteoarthrosis and internal derangement. *J Orofac Pain*. 1993;7(2):183-195.

45. Slade GD, Fillingim RB, Sanders AE, et al. Summary of findings from the OPPERA prospective cohort study of incidence of first-onset temporomandibular disorder: Implications and future directions. *J Pain*. 2013;14(12 Suppl):T116-24. doi: 10.1016/j.jpain.2013.09.010 [doi].

46. Dworkin RH, Turk DC, Wyrwich KW, et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. *J Pain*. 2008;9(2):105-121. doi: S1526-5900(07)00899-1 [pii].

47. Farrar JT, Young JP,Jr, LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. *Pain*. 2001;94(2):149-158. doi: S0304-3959(01)00349-9 [pii].

48. Moore A, Moore O, McQuay H, Gavaghan D. Deriving dichotomous outcome measures from continuous data in randomised controlled trials of analgesics: Use of pain intensity and visual analogue scales. *Pain*. 1997;69(3):311-315. doi: S0304-3959(96)03306-4 [pii].

49. Ingram M, Choi YH, Chiu CY, et al. Use of the minimal clinically important difference (mcid) for evaluating treatment outcomes with tmjmd patients: A preliminary study(). *J Appl Biobehav Res*. 2011;16(3-4):148-166. doi: 10.1111/j.1751-9861.2011.00068.x [doi].

50. Angst F, Aeschlimann A, Stucki G. Smallest detectable and minimal clinically important differences of rehabilitation intervention with their implications for required sample sizes using WOMAC and SF-36 quality of life measurement instruments in patients with osteoarthritis of the lower extremities. *Arthritis Rheum*. 2001;45(4):384-391. doi: 10.1002/1529-0131(200108)45:4<384::AID-ART352>3.0.CO;2-0 [doi].

51. Stewart AL, Greenfield S, Hays RD, et al. Functional status and well-being of patients with chronic conditions. results from the medical outcomes study. *JAMA*. 1989;262(7):907-913.

52. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: The remarkable universality of half a standard deviation. *Med Care*. 2003;41(5):582-592. doi: 10.1097/01.MLR.0000062554.74615.4C [doi].

53. Wyrwich KW, Nienaber NA, Tierney WM, Wolinsky FD. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. *Med Care*. 1999;37(5):469-478.

54. Ware JE,Jr, Bayliss MS, Rogers WH, Kosinski M, Tarlov AR. Differences in 4-year health outcomes for elderly and poor, chronically ill patients treated in HMO and fee-for-service systems. results from the medical outcomes study. *JAMA*. 1996;276(13):1039-1047.

55. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: A health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol*. 1988;15(12):1833-1840.

56. Roland M, Morris R. A study of the natural history of back pain. part I: Development of a reliable and sensitive measure of disability in low-back pain. *Spine (Phila Pa 1976)*. 1983;8(2):141-144.

57. Ohrbach R, Bair E, Fillingim RB, et al. Clinical orofacial characteristics associated with risk of first-onset TMD: The OPPERA prospective cohort study. *J Pain*. 2013;14(12 Suppl):T33-50. doi: 10.1016/j.jpain.2013.07.018 [doi].

58. Savitz DA. Confounding: Assess consequences of innacurate confounder measurement. In: *Interpreting epidemiologic evidence: Strategies for study design and analysis.* New York, NY: Oxford University Press; 2003:146-152.

59. Bossuyt PM, Reitsma JB, Bruns DE, et al. The STARD statement for reporting studies of diagnostic accuracy: Explanation and elaboration. *Ann Intern Med*. 2003;138(1):W1-12. doi: 200301070-00012-W1 [pii].

60. Underwood MR, Barnett AG, Vickers MR. Evaluation of two time-specific back pain outcome measures. *Spine (Phila Pa 1976)*. 1999;24(11):1104-1112.

61. Von Korff M. Assessment of chronic pain in epidemiological and health services research. In: Turk DC, Melzack R, eds. *Handbook of pain assessment.* 3rd ed. New York, NY: Guilford Press; 2011:455-471.

62. Stewart WF, Lipton RB, Simon D, Liberman J, Von Korff M. Validity of an illness severity measure for headache in a population sample of migraine sufferers. *Pain*. 1999;79(2-3):291-301.

63. ISOQOL - International Society of Quality of Life Research. What is health-related quality of life research. http://www.isoqol.org/about-isoqol/what-is-health-related-quality-of-life-research. Accessed 10/04, 2015.

64. Osoba D, King M. Meaningful differences. In: Fayers P, Hays RD, eds. *Assessing quality of life in clinical trials.* New York, NY: Oxford Press; 2005:244-257.

65. Feeny D, Furlong W, Mulhern RK, Barr RD, Hudson M. A framework for assessing health-related quality of life among children with cancer. *Int J Cancer Suppl*. 1999;12:2-9. doi: 10.1002/(SICI)1097-0215(1999)83:12+<2::AID-IJC2>3.0.CO;2-M [pii].

66. Cronbach LJ, Meehl PE. Construct validity in psychological tests. *Psychol Bull*. 1955;52(4):281-302.