

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ODONTOLOGIA DE BAURU

Everardo Napoleão Santana de Araújo Júnior

**Avaliação da confiabilidade interexaminador dos testes sensoriais
quantitativos em estruturas mastigatórias**

BAURU
2018

Avaliação da confiabilidade interexaminador dos testes sensoriais quantitativos em estruturas mastigatórias

Dissertação apresentada a Faculdade de Odontologia de Bauru da Universidade de São Paulo para obtenção do título de Mestre em Ciências no Programa de Ciências Odontológicas Aplicadas, na área de concentração Biologia Oral.

Orientador: Prof. Dr Leonardo Rigoldi Bonjardim

**BAURU
2018**

Ar15a Araújo Jr, Everardo
Avaliação da confiabilidade interexaminador dos testes sensoriais quantitativos em estruturas mastigatórias. / Everardo Napoleão Santana de Araújo Júnior. – Bauru, 2018.
59 p.: il. ; 31cm.

Dissertação (Mestrado) – Faculdade de Odontologia de Bauru. Universidade de São Paulo

Orientador: Prof. Dr. Leonardo Rigoldi Bonjardim

Autorizo, exclusivamente para fins acadêmicos e científicos, a reprodução total ou parcial desta dissertação/tese, por processos fotocopiadores e outros meios eletrônicos.

Assinatura:

Data:

Comitê de Ética da FOB-USP
Protocolo nº: 56433716.0.0000.5417
Data: 26/09/2016

(Cole a cópia de sua folha de aprovação aqui)

DEDICATÓRIA

Dedico a toda minha família, em especial a meu pai e minha mãe, que sempre me apoiaram a seguir essa longa trajetória de estudos.

AGRADECIMENTOS

Aos meus pais **Everardo** e **Silvana** por me ensinarem os valores fundamentais da vida me educarem e me amarem acima de qualquer circunstância. Em especial a minha mãe **Silvana** por ser um verdadeiro modelo de força e vontade de viver. Ao meu pai **Everardo** melhor e maior exemplo de dentista que tenho, por ser em quem eu me inspiro a me aperfeiçoar dia após dia. Ao meu irmão **Patrick** que sempre me apoia e incentiva. Aos meus irmãos pequenos **Maria** e **Davi** que amo e me encham de carinho sempre que estou com eles.

A minha esposa **Mariana**, por ser a melhor pessoa que eu poderia ter ao meu lado em todos os momentos, por me incentivar e motivar nesse trabalho, por entender minhas dificuldades e tentar sempre contribuir para supera-las. Por simplesmente ser quem é e estar ao meu lado incondicionalmente.

Ao meu querido Orientador **Prof. Dr. Leonardo Rigoldi Bonjardim**, por ser um ser humano impar, pelo acolhimento, por sempre ser compreensivo, parceiro e ensinar a melhorar cada vez mais em tudo que nos propomos a fazer não só durante a trajetória do mestrado, mas também na vida.

Ao meu amigo, colega e “Co-orientador” **Prof. Dr. Yuri Martins Costa**, que vibramos juntos felicidades e tristezas durante as pesquisas do laboratório, e por tornar a pesquisa da minha dissertação mais relevante com muitos direcionamentos e ensinamentos, além de toda ajuda e suporte que me deu em conjunto com meu orientador.

A todos participantes da pesquisa, sem eles nada seria possível.

A **CAPES** pelo apoio financeiro com a bolsa de mestrado fornecida.

A Faculdade de Odontologia de Bauru no nome da diretora **Profa Dra Maria Aparecida Andrade Moreira Machado** por toda a infraestrutura e acolhimento.

A todo o departamento de Ciências Biológicas no nome do chefe do departamento **Prof. Dr. Gustavo Pompermaier Garlet**.

A minha colega de pesquisa **Letícia Fiedler** que me ajudou a coletar os dados de toda a pesquisa.

A minha amiga **Fernanda Sampaio**, por me apresentar a DTM e me incentivar a fazer mestrado.

Ao IC **Paulo Jara** que me ajudou na tabulação de dados.

Aos meus colegas da graduação **Juliana, Mateus, Igor, Germana, Júlia, Yara, Brenda, Douglas, Saulo e Jardel** por me ajudarem e incentivarem durante toda a graduação e fora dela.

Aos meus amigos e parceiros da **Gracie Gym Zero 14 Brazilian Jiu Jitsu** que tornam a vida em Bauru mais fácil e divertida.

“Os que se encantam com a prática sem a ciência são como os timoneiros que entram no navio sem timão nem bússola, nunca tendo certeza do seu destino. “

Leonardo da Vinci

RESUMO

Avaliação da confiabilidade interexaminador dos testes sensoriais quantitativos em estruturas mastigatórias.

Este estudo estimou a confiabilidade interexaminador e a concordância da avaliação somatossensorial realizada na região do músculo masseter e da articulação temporomandibular (ATM) em um grupo de voluntários saudáveis do sexo feminino e masculino. Quarenta participantes saudáveis (20 homens e 20 mulheres) foram avaliados em duas sessões por dois examinadores diferentes. Os testes aplicados foram limiar de detecção ao frio (CDT, siglas em inglês), limiar de detecção ao calor (WDT), limiar sensorial térmico (TSL), limiar de dor ao frio (CPT), limiar de dor ao calor (HPT), limiar de detecção mecânica (MDT), limiar de dor mecânica (MPT), Razão da somação temporal (WUR) e o limiar de dor a pressão (PPT). ANOVA, coeficiente de correlação intraclasse (ICC) e erro padrão de mensuração (SEM) foram computados ($\alpha = 5\%$). O intervalo de confiança de 95% sem sobreposição (IC 95%) dos valores de ICC foi considerado significativamente diferente. Setenta e sete por cento dos valores de ICC de todas as medidas do QST variaram entre razoável a excelente (ICCs: 0,47-0,97). No entanto, a confiabilidade do WDT, TSL e HPT do masseter foi significativamente maior do que da ATM, enquanto que a confiabilidade de MDT da ATM foi maior do que do músculo masseter. Além disso, a seguinte combinação de teste/sítio apresentou os menores valores de ICC nas mulheres: HPT e MDT da ATM e MPT de ambos os sítios ATM e masseter. Para o erro padrão de mensuração, os maiores valores foram apresentados no CPT e MPT. A avaliação somatossensorial geral das estruturas de mastigatórias realizadas por dois examinadores pode ser considerada suficientemente confiável para discriminar os participantes. No entanto, as possíveis influências do sítio e do sexo nos parâmetros de reprodutibilidade devem ser levadas em consideração para uma interpretação e aplicação clínica apropriadas.

Palavras-chave: Limiar Sensorial. Articulação temporomandibular. Sistema estomatognático. Reprodutibilidade dos Testes

ABSTRACT

Evaluation of interrater reliability of quantitative sensory testing in the masticatory structures.

This study estimated the interrater reliability and agreement of the somatosensory assessment performed at masseter and temporomandibular joint (TMJ) region in a group of healthy female and male participants. Forty healthy participants (20 men and 20 men) were evaluated in two sessions by two different examiners. Cold detection threshold (CDT), warm detection threshold (WDT), thermal sensory limen (TSL), cold pain threshold (CPT), heat pain threshold (HPT), mechanical detection threshold (MDT), mechanical pain threshold (MPT), wind-up ratio (WUR) and pressure pain threshold (PPT) were assessed on the skin overlying masseter body and TMJ. Mixed ANOVA, Intraclass Correlation Coefficients (ICC) and standard error of measurement (SEM) were applied to the data ($\alpha=5\%$). Non-overlapping 95% confidence intervals (95% CI) of ICC values were considered significantly different. The ICC values of 77% of all QST measurements were considered fair to excellent (ICCs: 0.47-0.97). However, the reliability of WDT, TSL and HPT of masseter was significantly higher than TMJ, whereas the MDT reliability of TMJ was higher than masseter. In addition, the following combination of test/sites presented significantly lower ICCs for women: HPT, MDT of TMJ and MPT of both TMJ and masseter. Finally, the highest SEM values were presented for CPT and MPT. The overall somatosensory assessment of the masticatory structures performed by two examiners can be considered sufficiently reliable to discriminate participants. Nevertheless, possible site and sex influences on the reproducibility parameters should be take into account for an appropriate interpretation and clinical application.

Key words: Sensory Thresholds. Temporomandibular joint. Stomatognathic System. Reproducibility of Results

LISTA DE ILUSTRAÇÕES

FIGURAS

- Figura 1 - Perfil somatossensorial de indivíduos com diferentes quadros de dor baseado na transformação do QST em Z-escore..... 23

LISTA DE TABELAS

- Tabela 1 - Média, desvio padrão (SD) e intervalo de confiança de 95% (IC) da média dos valores absolutos do teste sensorial quantitativos (QST) avaliados pelos dois examinadores (dados das mulheres e dos homens combinados)..... 39
- Tabela 2 - Média, desvio padrão (SD) e intervalo de confiança de 95% (IC) da média dos valores absolutos do teste sensorial quantitativos (QST) avaliados pelos dois examinadores (dados das mulheres e dos homens separados)..... 40
- Tabela 3 - Análise de variância (ANOVA) considerando um fator dentro do grupo..... 39
- Tabela 4 - Confiabilidade relativa e valores de concordância da avaliação quantitativa dos testes sensoriais (QST) da articulação temporomandibular (ATM) e da região do masseter (dados masculinos e femininos combinados)..... 40
- Tabela 5 - Confiabilidade relativa e valores de concordância da avaliação de testes sensoriais quantitativos (QST) da articulação temporomandibular (ATM) e masseter (dados masculinos e femininos separados)..... 41

LISTA DE ABREVIATURA E SIGLAS

ATM	Articulação temporomandibular
QST	Teste Sensorial Quantitativo
CDT	Limiar de detecção ao frio
WDT	Limiar de detecção ao calor
TSL	Limiar sensorial térmico
CPT	Limiar de dor ao frio
HPT	Limiar de dor ao calor
MDT	Limiar de detecção mecânica
MPT	Limiar de dor mecânica
WUR	Razão da somação temporal
PPT	Limiar de dor a pressão
ICC	Índice de correlação intraclasse
SEM	Erro padrão de mensuração
SDC	Menor mudança detectável

SUMÁRIO

1	INTRODUÇÃO	15
2	REVISÃO DE LITERATURA	19
3	PROPOSIÇÃO	27
4	MATERIAL E MÉTODOS	28
4.1	AMOSTRA E ÉTICA	31
4.2	DESENHO DO ESTUDO	31
4.3	AVALIAÇÃO SOMATOSSENSORIAL	32
4.4	ESTATÍSTICA	34
5	RESULTADOS	37
6	DISCUSSÃO	45
7	CONCLUSÕES	51
	REFERÊNCIAS	55

1 Introdução

1 INTRODUÇÃO

A utilidade clínica e o protocolo de avaliação sensorial denominado de teste sensorial quantitativo (QST, na sigla em inglês) desenvolvido pela Rede Alemã de Investigação em Dor Neuropática (DFNS, na sigla em alemão) já está bem estabelecido na literatura científica, bem como sua aplicação clínica (Baron et al., 2017; von Hehn et al., 2012). O entendimento do tratamento das dores de origem neuropática com base em perfis somatossensoriais ampliou a aplicação dos QSTs a outras condições dolorosas crônicas, por exemplo, distúrbios musculoesqueléticos que abrangem as disfunções temporomandibulares (DTMs), que é um termo coletivo que engloba condições clínicas que afetam as estruturas musculoesqueléticas do sistema mastigatório (de Leeuw, 2010) e podem ser associados com sinais de ganho somatossensorial para dor ao calor (hiperalgesia térmica), ao toque (alodinia), à pressão (hiperalgesia mecânica) e um aumento da somação temporal (fenômeno de *wind up*) (Kothari et al., 2015; Pfau et al., 2009; Yang et al., 2016). Além disso, já é possível distinguir subgrupos de pacientes com DTM dolorosa baseado em diferentes perfis somatossensoriais (Bair et al., 2016; Pfau et al., 2009). No entanto, a aplicabilidade clínica da avaliação somatossensorial nos distúrbios musculoesqueléticos não é totalmente clara. Portanto, uma investigação abrangente sobre as propriedades de medidas da bateria do QST em áreas anatômicas importantes do sistema mastigatório, como músculos da mastigação em conjunto com a articulação temporomandibular (ATM), poderia contribuir para elucidar a interpretação dos resultados do QST na dor orofacial de origem musculoesquelética.

Embora haja uma quantidade razoável de evidências relacionadas à confiabilidade das aplicações do QST nas regiões espinhais e intraorais (Andersen et al., 2015; Baad-Hansen et al., 2013; Geber et al., 2011; Lindgren et al., 2016; Marcuzzi et al., 2017; Nothnagel et al., 2017; Pigg et al., 2010) a reprodutibilidade da avaliação somatossensorial nas estruturas mastigatórias não é adequadamente conhecida. Até onde se sabe, apenas a confiabilidade relativa do teste-reteste do QST mecânico foi estimada para as estruturas mastigatórias e os resultados apresentaram valores de confiabilidade aceitáveis em participantes saudáveis (Costa et al., 2017). Além disso, observando que as respostas somatossensoriais diferem entre as várias regiões anatômicas (Backonja et al.,

2009; Backonja et al., 2013; Rolke et al., 2006), diferenças específicas entre sítios anatômicos nos parâmetros de reprodutibilidade não seriam surpreendentes. Do mesmo modo, também não é amplamente conhecida se as eventuais diferenças na sensibilidade somatossensorial entre mulheres e homens poderiam influenciar a confiabilidade do QST (Fillingim et al 2017; Riley et al., 1998).

Nessa direção, o presente estudo estimou a confiabilidade interexaminador e a concordância da avaliação somatossensorial realizada na região do músculo masseter e da articulação temporomandibular em um grupo de voluntários saudáveis de ambos os sexos.

2 Revisão de Literatura

2 REVISÃO DE LITERATURA

A Medicina preconizou durante muitos anos o uso do modelo biomédico no entendimento das doenças. Esse modelo era pautado em ver o homem como uma máquina e focava apenas em processos físicos da doença, como por exemplo a patologia e a bioquímica. A partir do desenvolvimento da psicofísica com o estudo das relações entre sensações subjetivas e estímulos físicos, tornou-se possível o melhor entendimento do corpo humano, em particular das percepções humanas (de Leeuw, 2010).

A ciência dos fenômenos sensoriais surgiu com o objetivo de criar uma maneira capaz de quantificar as sensações físicas e psíquicas. Com o avanço de recursos e de tecnologia, buscou-se ferramentas com o melhor custo-benefício que fossem capazes de quantificar as sensações (Backonja et al., 2013).

Para “traduzir” essas sensações e trazê-las para a clínica de diagnóstico e tratamento da dor, o mais importante seria caracterizar o fenótipo somatossensorial do indivíduo com a maior precisão possível, utilizando-se de uma avaliação completa e sistematizada do padrão somatossensorial. Nessa direção, a Rede Alemã de Pesquisa em Dor Neuropática (DFNS, na sigla em alemão) estabeleceu um protocolo padronizado capaz de traçar esse perfil somatossensorial. Esse conjunto de testes psicofísicos, conhecidos como teste sensorial quantitativo (QST, na sigla em inglês), utiliza-se de diferentes modalidades estímulos sensoriais (térmico, mecânico e doloroso) para verificar o ganho e a perda de funções somatossensoriais e assim determinar diferentes tipos de distúrbios, analisando desde a sensibilidade cutânea até dor profunda (Rolke et al., 2006).

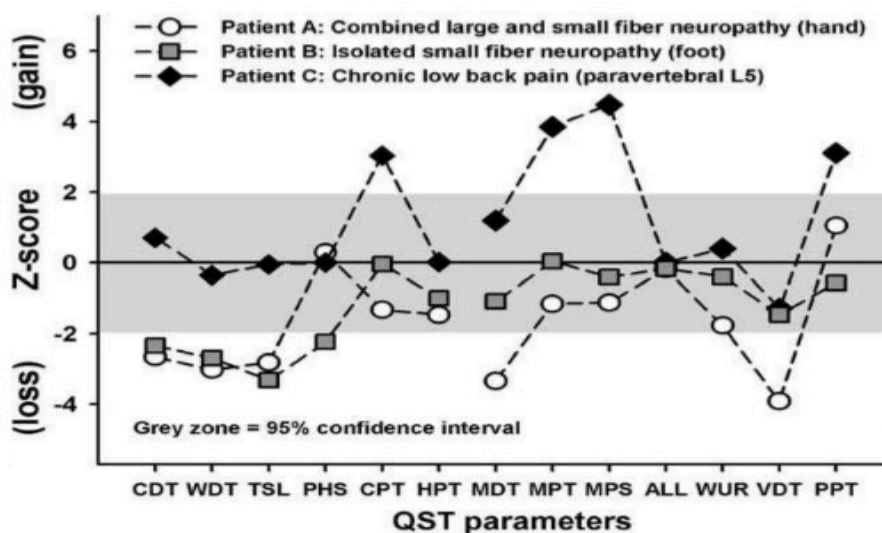
O protocolo de QST compreende uma bateria de 13 testes sensoriais divididos sequencialmente em: 1) Térmico que compreende a mensuração dos limiares de detecção e de dor para estímulos de frio ou calor (mediadas por fibras C e A-delta): limiar de detecção de frio (CDT, sigla em inglês); limiar de detecção de calor (WDT, sigla em inglês); limiar sensorial térmico alternados para estímulos de frio e calor (TSL, sigla em inglês); número de sensações de calor paradoxais durante o teste TSL (PHS, sigla em inglês); limiar de dor ao frio (CPT, sigla em inglês); limiar de dor ao calor (HPT, sigla em inglês). 2) Mecânico que compreende os testes de limiar de detecção mecânica (MDT, sigla em inglês) e de vibração (VDT, sigla em inglês) os quais mensuram anormalidades sensoriais para a função de fibra

nervosas do tipo A-beta em tecidos cutâneos e mais profundos. 3) Doloroso que compreende os testes de limiar de dor mecânica (MPT, sigla em inglês) que mensura fenômenos de ganho ou perda sensorial mediados por fibras A-delta utilizando estímulos do tipo agulhada; Sensibilidade à dor mecânica (MPS, sigla em inglês) para estímulos de agulhada usando filamentos com pressão calibrada; Alodínia mecânica dinâmica (DMA, sigla em inglês), avaliam a sensibilidade à dor mediada por estímulos mecânicos inócuos (algodão, ponta de algodão e escova); Razão da somação temporal (WUR, sigla em inglês) que mensura a razão da intensidade da dor numa escala de 0 a 100 entre um único estímulo doloroso (a) e uma série de dez estímulos dolorosos (b), podendo-se calcular essa razão (WUR) como a relação b/a ; Limiar de dor à pressão (PPT, sigla em inglês) é o único teste que mensura a sensibilidade à dor profunda, utilizando um algômetro ou dinamômetro de pressão, provavelmente mediado pelas fibras do tipo C e A-delta (Rolke et al., 2006).

Para uma melhor interpretação individualizada dos dados obtidos pelo QST é preciso transformar os valores de cada teste somatossensorial em um Z-escore para produzir um perfil somatossensorial relacionado a cada QST. Essa manipulação matemática, que leva em consideração valores normativos de um grupo controle, ajuda a determinar a relevância dos resultados de cada paciente. Assim, a partir disso, se analisa o perfil do QST e a pontuação-z acima de "0" indica um ganho de função à uma ou mais das 13 modalidades sensoriais testadas e uma pontuação-z abaixo de "0" indica uma perda de função referente a uma menor sensibilidade do paciente a um ou mais dos estímulos testados. Os Z-escores que representam um intervalo de normalidade podem variar de $\pm 1,96$ e representam o intervalo que inclui 95% dos valores de referência, sendo que quaisquer pontuações fora desse intervalo podem ser consideradas anormalidades somatossensoriais (Rolke et al., 2006).

A figura abaixo foi retirada do artigo de Rolke et al. (2006) e exemplifica alguns resultados do QST baseado na transformação em Z-escore de alguns pacientes selecionados. O paciente A apresenta o perfil somatossensorial que mostra perda de função sensorial para estímulos térmicos e mecânicos. Pode-se observar na figura, os estímulos mecânicos MDT e VDT e térmicos CDT, WDT e TSL abaixo do intervalo de confiança de 95% da distribuição dos valores de indivíduos saudáveis, representado pela zona cinza. O paciente B apresenta-se

também com anormalidade sensorial, sendo possível observar os parâmetros CDT, WDT, TSL, PHS fora do intervalo normal, representado pela zona cinza do gráfico. Por fim, o paciente C mostra o perfil do QST de uma mulher de 45 anos de idade com dor lombar crônica e diferente dos pacientes A e B descreve um perfil de ganho da função sensorial para os parâmetros MPT, MPS e PPT.



Fonte: Rolke, R., Baron, R., Maier, C. A., Tölle, T. R., Treede, R. D., Beyer, A., ... & Braune, S. (2006). Quantitative sensory testing in the German Research Network on Neuropathic Pain (DFNS): standardized protocol and reference values. *Pain*, 123(3), 231-243.

Figura 1 – Perfil somatossensorial de indivíduos com diferentes quadros de dor baseado na transformação do QST em Z-escore.

De maneira sistematizada, os testes sensoriais quantitativos surgiram inicialmente na área médica, mais especificamente para ser utilizado na compreensão das dores neuropáticas (Rolke et al., 2006). De maneira sequencial, em 2011, foi publicado um guia sobre a avaliação da função somatossensorial na região orofacial, em termos de técnicas e desempenho dos testes com o intuito de fornecer recomendações do uso do QST na região orofacial intraoral (Svenson et al., 2011). Com base na literatura disponível, os autores verificaram que a bateria de testes proposta pelo DFNS parecia apropriada para o estudo da função somatossensorial na região orofacial, pois baseia-se em uma grande variedade de avaliações quantitativas de todas as modalidades somatossensoriais cutâneas e de tecido profundo. Ainda, os autores concluíram haver necessidade de padronização dos testes somatossensoriais a fim de melhorar a precisão do diagnóstico e facilitar a compreensão dos mecanismos neurais e alterações somatossensoriais em

diferentes condições de dor orofacial o que poderá ajudar a orientar o tratamento (Svesson et al., 2011).

Os estudos têm utilizado a determinação do perfil somatossensorial orofacial, por meio do QST, com diferentes objetivos. Por exemplo, o QST foi utilizado na região trigeminal para avaliar anormalidades sensoriais em pacientes antes e após cirurgia ortognática e verificou-se que uma grande parte dos pacientes estudados após a cirurgia eram hipersensíveis aos estímulos mecânicos nas áreas trigeminais e extratrigeminais, indicando presença de sensibilização regional e central (Baad-Hansen et al., 2010).

Esses mesmos testes sensoriais quantitativos também foram utilizados em um estudo que avaliou anormalidades sensoriais em várias dores neuropáticas, incluindo neuralgia do trigêmeo, onde pode-se perceber anormalidade sensoriais em todas as síndromes neurológicas avaliadas (Maier et al., 2010).

O QST também tem sido usado para mensurar anomalias somatossensoriais na região intraoral, especialmente em pacientes com odontalgia atípica. Um estudo encontrou que pacientes que apresentavam esse diagnóstico clínico, quando eram avaliados por meio do QST, frequentemente apresentavam um ganho somatossensorial em relação aos estímulos mecânicos e de dor ao frio, apresentando também em alguns casos, uma resposta de perda somatossensorial em relação aos estímulos de detecção do frio e de detecção mecânica (Baad-Hansen et al., 2013).

Com o avanço do QST na região orofacial começaram a ser publicados estudos buscando traçar um perfil somatossensorial de pacientes com dores de origem musculoesquelética, como a DTM (Pfau et al., 2009; Greenspan et al., 2013). Paralelo a isso, o QST foi referenciado na última atualização dos Critérios de Diagnóstico para Disfunção Temporomandibular (DC/TMD), como um possível e importante biomarcador a ser incluído em um futuro eixo III desses critérios, mostrando a importância da determinação do perfil somatossensorial nos quadros de DTM (Schiffman et al., 2014).

Alguns estudos recentes enfatizam o seu potencial para um prognóstico mais preciso baseado no perfil somatossensorial em pacientes com DTM (Kothari et al 2015; Hilgenberg-Sydney et al., 2016; Yang et al., 2016). Um estudo que utilizou QST para avaliar se a função somatossensorial na ATM e a modulação da dor condicionada (CPM) diferia entre pacientes com DTM dolorosa e controle saudáveis

concluiu que anormalidades somatossensoriais eram comumente encontradas em pacientes com DTM dolorosa, porém os efeitos de CPM foram semelhantes em pacientes com DTM dolorosa e controles saudáveis (Kothari et al., 2015).

Outro estudo avaliou pacientes com diferentes tipos de dores crônicas como dor miofascial dos músculos mastigatórios, fibromialgia, dor de cabeça crônica diária e voluntários saudáveis encontrando que indivíduos sintomáticos apresentaram valores mais baixos para o limiar de dor mecânica e para a tolerância à dor isquêmica do que voluntários saudáveis e que a capacidade de ativar o mecanismo de modulação endógena é prejudicada em mulheres com fibromialgia e dor miofascial. (Hilgenberg-Sydney et al., 2016). Um último estudo avaliou o perfil somatossensorial na região facial mais dolorosa, no lado contralateral e no dorso da mão direita. Os autores encontraram que anormalidades somatossensoriais foram comumente detectadas em chineses com DTM dolorosa dentro e fora da região que apresentava algum tipo de dor, indicando fortemente distúrbios no processamento central da dor (Yang et al., 2016).

Entretanto, avaliar a reprodutibilidade qualquer tipo de teste na área médica é crucial para o desenvolvimento de pesquisas e aplicabilidade clínica. De uma forma bem simples, a reprodutibilidade diz respeito ao grau em que as medidas repetidas proporcionam resultados comparáveis e semelhantes. A reprodutibilidade pode ser expressa em parâmetros de confiabilidade e concordância. A concordância (agreement, termo em inglês) significa o quão exata ela é entre medidas repetidas e pode ser medida pelo erro padrão de mensuração (SEM, sigla em inglês), que avalia exatamente o quão próximo são os valores para medições repetidas. Já a confiabilidade refere-se à quão bem os sujeitos podem ser distinguidos um do outro apesar dos erros de medição. Nesse caso, o erro de medição é conjugado com a variabilidade entre sujeitos de estudo (de Vet et al., 2006). A confiabilidade de variáveis quantitativas (reliability, termo em inglês) pode ser medida pelo coeficiente de correlação intraclassa (ICC, sigla em inglês).

Ainda, pelo QST ser um teste psicofísico, que depende da resposta do indivíduo, a avaliação de sua confiabilidade se torna ainda mais imprescindível. Trabalhos que avaliam a confiabilidade e outras propriedades métricas do QST aplicado na região orofacial ainda são incipientes. Um estudo avaliou as diferenças das respostas do QST em relação a gênero e sítios, utilizando o protocolo completo de QST nas saídas nervosas extraorais e encontrou que não havia diferença quando

os sítios eram avaliados, mas que houve diferença significativa de gênero, com limiar mais baixo (maior sensibilidade) em mulheres (Mato et al., 2011).

Um estudo de confiabilidade do QST na região orofacial, que foi realizado pelo nosso grupo de pesquisa, estimou a confiabilidade teste-reteste da avaliação mecânica somatossensorial na região do temporal anterior, masseter e articulação temporomandibular (ATM). Os principais achados desse estudo foram que os valores gerais de confiabilidade teste-reteste da avaliação somatossensorial mecânica do temporal anterior, masseter e ATM são aceitáveis para aplicação clínica (Costa, et al., 2017).

Um outro estudo que avaliou a confiabilidade inter e intraexaminador do protocolo completo do QST na região intraoral e na área da bochecha observou que a confiabilidade pode variar de acordo com o sítio e com o tipo de teste de QST, mas de uma maneira geral a confiabilidade foi aceitável para todos os testes (Pigg et al., 2010). Por fim um estudo, do mesmo grupo de pesquisadores, avaliou a confiabilidade teste-reteste e interexaminador do QST na região intraoral e encontrou que a confiabilidade pode variar de razoável a excelente de acordo com o tipo de teste aplicado, concordando com o estudo anterior (Baad-Hansen et al., 2015). Portanto, para o nosso conhecimento, ainda não existem trabalhos que avaliaram a confiabilidade interexaminador do QST em estruturas mastigatórias de interesse para a disfunção temporomandibular (DTM).

3 Proposição

3 PROPOSIÇÃO

O objetivo deste estudo foi estimar a confiabilidade interexaminador e a concordância da avaliação somatossensorial realizada em masseter e ATM em um grupo de participantes saudáveis do sexo feminino e masculino.

A hipótese geral foi: o sitio do teste e o sexo dos participantes influenciariam significativamente as estimativas de confiabilidade do QST.

Material e Métodos

4 MATERIAL E MÉTODOS

4.1 *Amostra e Ética*

Quarenta participantes saudáveis (20 homens e 20 mulheres) foram recrutados da comunidade de alunos de graduação, pós-graduação e funcionários da Faculdade de Odontologia de Bauru, ou seja, método de amostragem de conveniência. Os critérios de inclusão foram: idade > 18 anos, boa saúde sistêmica sem queixas de dor orofacial nos últimos 30 dias ou distúrbios de dor crônica, ausência de doença dentária ou médica, sem ingestão regular de medicação como antidepressivos, anticonvulsivantes ou anti-inflamatórios não esteroides e ausência de transtornos psiquiátricos ou de personalidade. Uma entrevista / anamnese detalhada foi realizada para avaliar os critérios de inclusão e exclusão acima detalhados.

Este estudo foi realizado de acordo com a Declaração de Helsink II e teve a aprovação do Comitê de Ética em Pesquisa com Seres Humanos da Faculdade de Odontologia de Bauru, Universidade de São Paulo (CAAE: 56433716.0.0000.5417). Todos os participantes deram o seu consentimento voluntário após uma explicação completa de todos os procedimentos.

4.2 *Desenho do Estudo*

Todos os participantes foram avaliados em duas sessões (avaliação inicial e uma semana depois) por dois examinadores (examinador 1 - EAJ e examinador 2 - LF), a fim de se estimar a confiabilidade relativa interexaminador e a concordância. Nenhum dos examinadores apresentavam experiência prévia com avaliação somatossensorial, portanto cada um dos examinadores foi treinado por um especialista certificado em QST pelo DNFS (YC). A conveniência orientou a escolha dos examinadores e ambos estavam conscientes de que seus resultados seriam comparados, ou seja, eles tinham conhecimento prévio de que suas avaliações seriam usadas para estimar a confiabilidade interexaminador. As mensurações foram conduzidas de forma independente e nenhum dos examinadores estava ciente dos resultados do outro examinador.

O perfil somatossensorial foi avaliado na área do corpo do músculo masseter e na articulação tempomandibular (ATM) no lado correspondente à mão dominante do voluntário. A ordem dos testes somatossensoriais seguiu a descrição original das

diretrizes do DFNS (Rolke et al., 2006). Cada bateria de testes em cada sítio levou aproximadamente 20 minutos para ser realizada. Finalmente, a ordem dos sítios foi aleatorizada e foi mantido a mesma para ambas as sessões, também sendo randomizada a ordem do turno do examinador. Sendo assim, possíveis erros sistemáticos devido a um examinador ser sempre o primeiro a examinar os participantes poderiam ser evitados.

4.3 *Avaliação Somatossensorial*

Uma avaliação somatossensorial quantitativa padronizada foi realizada de acordo com as diretrizes do DFNS (Rolke et al., 2006). Foram avaliados nove dos 13 parâmetros recomendados: limiar de detecção do frio (CDT, todas as siglas em inglês), limiar de detecção do calor (WDT), limiar térmico sensorial (TSL), limiar de dor ao frio (CPT), limiar de dor ao calor (HPT), limiar de detecção mecânica (MDT), limiar de dor mecânica (MPT), razão da somação temporal (WUR) e limiar de dor à pressão (PPT).

Os testes térmicos *CDT*, *WDT*, *TSL*, *CPT* e *HPT* foram realizados com o auxílio de estimulador térmico computadorizado TSA 2001-II (MEDOC, Ramat Yishai, Israel). Para a determinação do CDT e do WDT, o probe/sensor térmico do TSA 2001 – II foi colocado sobre o sítio determinado (área de contato do probe foi de 2,56 cm²) e foi solicitado ao voluntário que interrompesse o teste quando percebesse a primeira sensação de frio para a avaliação do CDT e a primeira sensação de calor para a avaliação do WDT. Da mesma forma, foi solicitado ao voluntário que interrompesse o teste quando a primeira sensação de dor ao frio fosse percebida para a avaliação do CPT e a primeira sensação de dor ao calor para a avaliação do HPT. Foram feitas três medições e a média aritmética foi considerada para determinação dos do CDT, WDT, CPT e HPT e como descrito acima, a determinação dos limiares foi feita pelo método dos limites. A temperatura da linha de base foi de 32°C e foram aplicados estímulos em rampa (1°C/s) que foram finalizados quando o sujeito pressionou um botão ligado ao equipamento TSA 2001-II (Rolke et al., 2006). Além disso, estímulos frios e quentes foram aplicados para determinar o TSL.

A medição dos testes mecânicos MDT e MPT foi realizada por meio de um conjunto padronizado de monofilamentos de Semmes-Weinstein (Touch-Test™, North Coast Medical Inc., Gilroy, EUA), que aplicam forças entre 0,008

g/mm² a 300 g/mm². Os participantes foram instruídos a relatar a primeira sensação de toque para a avaliação de MDT e a relatar a primeira sensação dolorosa de picada para a avaliação de MPT. Os filamentos foram aplicados em uma posição vertical e perpendicular ao local de exame até sua flexão e o tempo de contato foi de aproximadamente 2s. Para ambos os testes mecânicos (MDT e MPT), foi utilizado o método de limites modificado. Assim para a determinação do MDT, partiu-se do filamento menos calibroso e foram aplicados sequencialmente filamentos cada vez mais calibrosos até que o voluntário reportasse verbalmente sentir um leve toque como instruído no início teste. Isso foi considerado como um estímulo positivo (+) ou supra limiar. Após esse relato positivo, se inverteu a ordem e seguiu-se para o próximo filamento com valor mais baixo, até que o voluntário não sentisse o leve toque. Isso foi considerado como um estímulo negativo (-) ou sub limiar. Essa sequência foi repetida dez vezes, sendo obtidos cinco estímulos negativos e cinco positivos, sendo o único limiar calculado pela média geométrica desses dez limiares (Rolke et al., 2006). Para a determinação do MPT foi seguida as mesmas normas do MDT descrito anteriormente, com a diferença da instrução do voluntário que deveria reportar quando o filamento evocasse uma sensação dolorosa correspondente a uma “agulhada, alfinetada ou picada” e não mais a sensação de toque leve. Assim, foram determinados cinco estímulos negativos que referiram a ausência de sensação dolorosa e cinco estímulos positivos em que o filamento determinava uma sensação de dor, sendo o único limiar calculado pela média geométrica desses dez limiares (Rolke et al., 2006).

Para o WUR os monofilamentos de Semmes-Weinstein também foram escolhidos. Esses filamentos foram aplicados com uma força em que o participante percebesse o monofilamento como ligeiramente doloroso e indicava uma resposta numérica maior que 0 indicando a intensidade da dor. A intensidade percebida de uma única alfinetada era então comparada com uma série de 10 estímulos repetitivos de alfinetada da mesma intensidade física (aplicado dentro de uma área de 1 cm² numa taxa de um por segundo). O participante foi convidado a relatar a intensidade da dor que representa o estímulo único e a média estimada em todo o conjunto de 10 estímulos usando uma escala de classificação numérica de 0-100. Todo o procedimento foi repetido três vezes e a razão da somação temporal de dor (WUR) foi calculado pela média da intensidade da dor das três

séries de 10 estímulos repetidos divididas pela média da intensidade da dor dos três estímulos únicos (Rolke et al., 2006).

A medida de PPT foi realizada através de um dinamômetro digital (Kratos®, Cotia, Brasil) contendo uma haste com uma ponta plana e circular de 1 cm² no final dela. Esse dinamômetro foi utilizado para aplicar pressão sobre o corpo do músculo masseter e sobre a ATM, com uma taxa de aplicação de aproximadamente 0,5 kgf/cm²/seg. Os participantes foram instruídos a pressionar um botão de parada do teste na primeira sensação dolorosa adicionada à sensação de pressão. Enfatizou-se que o objetivo era medir a quantidade mínima de pressão na primeira percepção da dor e não a tolerância à dor. O PPT foi determinado como a média aritmética de três medidas (Rolke et al., 2006).

4.4 Estatísticas

Os valores das variáveis sensoriais foram expressos como média, desvio padrão (SD, sigla em inglês) e intervalo de confiança de 95% (IC 95%) da média. A distribuição dos valores absolutos do QST foi avaliada pelo teste de Kolmogorov-Smirnov e uma transformação de log₁₀ foi realizada quando os resultados foram significativos, considerando um nível de significância de 5% ($p < 0,050$). Assim, os seguintes parâmetros foram transformados: WDT, CPT, MDT, MPT, WUR e PPT. Depois disso, uma análise de variância (ANOVA) foi computada para comparar os valores de QST da avaliação do examinador 1 considerando um fator intrasujeito, i.e., sítio (2 níveis), e um fator intersujeito, i.e., sexo (2 níveis). Quando apropriado, análises *post-hoc* foram realizadas usando o Tukey Honestly Statistical Difference (HSD). O nível de significância foi estabelecido em 5% ($p = 0,050$).

A confiabilidade interexaminador (confiabilidade relativa) foi estimada usando o coeficiente de correlação intraclassa (ICC, sigla em inglês) e seu IC de 95% (Stratford et al., 1997). Os ICCs foram obtidos usando uma única medida com base em ANOVA e modelo aleatório de duas vias. A magnitude do ICC foi classificada como ruim ($< 0,4$), razoável (0,4 - 0,59), bom (0,6 - 0,75) ou excelente ($> 0,75$) (Stratford et al., 1997). Os 95% IC não sobrepostos foram considerados significativamente diferentes.

A concordância (confiabilidade absoluta) foi estimada usando o erro padrão de mensuração (SEM, sigla em inglês), que foi calculado com a seguinte fórmula:

SEM = $\sqrt{(\sigma^2_{ex} + \sigma^2_{residual})}$, onde σ^2_{ex} indica a variância devido a diferenças sistemáticas entre os examinadores e o $\sigma^2_{residual}$ indica a variância do erro aleatório, ou seja, a interação das diferenças sistemáticas da combinação entre participantes e examinadores (de Vet et al., 2006; Stratford et al., 1997).

Finalmente, a menor mudança detectável (SDC, sigla em inglês), ou seja, a menor quantidade de alteração no escore que é maior que o erro de mensuração, foi calculada com a seguinte fórmula: SDC = $1.96 * \sqrt{2} * SEM$ [9; 27]. Todos os testes foram realizados utilizando o Pacote Estatístico para Ciências Sociais (SPSS, sigla em inglês) v.18.0 (IBM Corp., EUA) e STATISTICA, v 12 (StatSoft Inc., EUA).

5 Resultados

5 RESULTADOS

Vinte mulheres com idade média de 24,6 (3,9) anos e 20 homens com idade de 26,0 (4,5) anos foram avaliados por dois examinadores em duas sessões e os valores absolutos do QST dos dois examinadores estão descritos nas tabelas 1 e 2. Somente os valores de MDT e WUR foram semelhantes entre sítios e sexo sem efeitos principais ou interações significativas (ANOVA: $F < 3,09$, $p > 0,086$) (Tabela 3). No entanto, foi observada diferença significativa entre sítios para o CDT, WDT, TSL e HPT, onde a ATM apresentou menores limiares quando comparados com o masseter (Tabela 3). Além disso, diferença significativa foi encontrada para o sexo nos testes CPT, HPT, MPT e PPT, sendo as mulheres mais sensíveis do que os homens (Tabela 3). Finalmente, uma diferença significativa entre sítio e sexo foi observada para o PPT, onde o masseter dos homens apresentou limiares mais altos do que a ATM (Tukey: $p < 0,009$) e masseter (Tukey: $p < 0,001$) das mulheres (tabela 3).

Tabela 1. Média, desvio padrão (SD, sigla em inglês) e intervalo de confiança de 95% (IC) da média dos valores absolutos do teste sensorial quantitativo (QST) avaliados pelos dois examinadores (dados das mulheres e dos homens combinados).

	Examinador 1		Examinador 2	
	Média (SD)	Média – 95% IC	Média (SD)	Média – 95% IC
ATM				
CDT °C	28,71(1,60)	28,20 – 29,22	28,16 (1,71)	27,61 – 28,71
WDT °C	36,01(1,86)	35,42 – 36,61	36,39 (2,59)	35,56 – 37,22
TSL °C	6,81(2,56)	5,99 – 7,63	6,83 (3,42)	5,74 – 7,93
CPT °C	13,82 (10,38)	10,50 – 17,14	13,72 (9,92)	10,55 – 16,90
HPT °C	43,58 (3,73)	42,39 – 44,78	43,68 (4,45)	42,25 – 45,10
MDT (g/mm ²)	0,15 (0,73)	-0,08 – 0,39	0,24 (0,79)	-0,01 – 0,49
MPT (g/mm ²)	53,96 (68,0)	32,21 – 75,71	50,22 (64,03)	29,74 – 70,70
WUR (NRS)	1,69 (0,81)	1,43 – 1,95	2,05 (0,95)	1,74 – 2,35
PPT (kgf/cm ²)	1,53 (0,45)	1,38 – 1,67	1,53 (0,46)	1,38 – 1,68
Masseter				
CDT °C	27,85 (2,04)	27,20 – 28,50	27,51 (1,96)	26,88 – 28,14
WDT °C	38,24 (3,75)	37,04 – 39,44	38,72 (3,85)	37,49 – 39,95
TSL °C	10,23 (5,20)	8,57 – 11,90	9,81 (4,47)	8,38 – 11,24
CPT °C	13,72 (9,66)	10,63 – 16,81	14,25 (9,66)	11,16 – 17,34
HPT °C	45,15 (3,76)	43,95 – 46,35	45,33 (3,50)	44,21 – 46,45
MDT (g/mm ²)	0,07 (0,11)	0,04 – 0,11	0,17 (0,36)	0,06 – 0,29
MPT (g/mm ²)	45,35 (59,15)	26,43 – 64,27	54,16 (71,00)	31,45 – 76,87
WUR (NRS)	1,70 (0,69)	1,48 – 1,92	2,17 (1,18)	1,79 – 2,55
PPT (kgf/cm ²)	1,50 (0,51)	1,33 – 1,66	1,55 (0,48)	1,39 – 1,70

Limite de detecção do frio (CDT), limiar de detecção do calor (WDT), limiar sensorial térmico (TSL), limiar de dor ao frio (CPT), limiar de dor ao calor (HPT), limiar de detecção mecânica (MDT), limiar de dor mecânica (MPT) razão da somação temporal (WUR) e limiar de dor a pressão (PPT). NRS = escala de classificação numérica.

Tabela 2. Média, desvio padrão (SD, sigla em inglês) e intervalo de confiança de 95% (IC) da média dos valores absolutos do teste sensorial quantitativo (QST) avaliados pelos dois examinadores (dados de mulheres e homens separados).

	Examinador 1		Examinador 2	
	Média (SD)	Média – 95% CI	Média (SD)	Média – 95% CI
Mulheres				
ATM				
CDT °C	28,57 (1,83)	27,71 – 29,43	28,15 (1,19)	27,59 – 28,71
WDT °C	36,17 (1,89)	35,28 – 37,06	36,54 (2,86)	35,20 – 37,88
TSL °C	7,03 (2,78)	5,72 – 8,33	7,03 (3,22)	5,52 – 8,54
CPT °C	18,06 (9,22)	13,74 – 22,38	17,03 (8,34)	13,12 – 20,93
HPT °C	42,20 (2,76)	40,91 – 43,49	42,66 (4,47)	40,57 – 44,75
MDT (g/mm ²)	0,04 (0,07)	0,01 – 0,08	0,16 (0,23)	0,05 – 0,27
MPT (g/mm ²)	26,11 (31,69)	11,28 – 40,94	15,81 (12,24)	10,08 – 21,54
WUR (NRS)	1,47 (0,47)	1,25 – 1,69	2,10 (1,11)	1,57 – 2,62
PPT (kgf/cm ²)	1,33 (0,36)	1,15 – 1,50	1,24 (0,30)	1,10 – 1,38
Masseter				
CDT °C	27,49 (2,19)	26,46 – 28,52	27,50 (2,08)	26,52 – 28,48
WDT °C	38,03 (3,63)	36,33 – 39,73	37,95 (4,09)	36,04 – 39,87
TSL °C	10,64 (6,21)	7,73 – 13,55	9,88 (4,22)	7,90 – 11,85
CPT °C	16,53 (9,28)	12,18 – 20,88	18,22 (8,71)	14,14 – 22,30
HPT °C	43,73 (4,12)	41,80 – 45,66	44,31 (3,64)	42,60 – 46,01
MDT (g/mm ²)	0,05 (0,07)	0,01 – 0,08	0,15 (0,31)	0,00 – 0,30
MPT (g/mm ²)	20,79 (26,45)	8,41 – 33,17	25,20 (25,98)	13,04 – 37,36
WUR (NRS)	1,52 (0,48)	1,29 – 1,75	2,14 (1,25)	1,55 – 2,73
PPT (kgf/cm ²)	1,20 (0,38)	1,02 – 1,38	1,29 (0,41)	1,09 – 1,48
Homens				
ATM				
CDT °C	28,85 (1,36)	28,21 – 29,49	28,17 (2,14)	27,17 – 29,17
WDT °C	35,86 (1,86)	34,99 – 36,73	36,25 (2,36)	35,14 – 37,35
TSL °C	6,59 (2,37)	5,48 – 7,70	6,64 (3,68)	4,91 – 8,36
CPT °C	9,58 (9,90)	4,94 – 14,21	10,42 (10,47)	5,52 – 15,33
HPT °C	44,97 (4,11)	43,04 – 46,89	44,70 (4,31)	42,67 – 46,71
MDT (g/mm ²)	0,26 (1,04)	-0,22 – 0,75	0,32 (1,11)	-0,19 – 0,84
MPT (g/mm ²)	81,82 (82,81)	43,05 – 46,89	84,64 (74,98)	49,07 – 120,20
WUR (NRS)	1,91 (1,01)	1,437 – 2,38	2,00 (0,78)	1,63 – 2,36
PPT (kgf/cm ²)	1,74 (0,45)	1,52 – 1,94	1,82 (0,41)	1,62 – 2,01
Masseter				
CDT °C	28,21 (1,87)	27,33 – 29,09	27,52 (1,88)	26,64 – 28,40
WDT °C	38,45 (3,95)	36,60 – 40,29	39,48 (3,52)	37,83 – 41,13
TSL °C	9,83 (4,07)	7,92 – 11,73	9,75 (4,82)	7,49 – 12,01
CPT °C	10,91 (9,41)	6,50 – 15,32	10,28 (9,08)	6,03 – 14,53
HPT °C	46,57 (2,79)	45,26 – 47,87	46,36 (3,11)	44,90 – 47,82
MDT (g/mm ²)	0,10 (0,14)	0,03 – 0,17	0,19 (0,40)	0,00 – 0,38
MPT (g/mm ²)	69,92 (72,20)	36,12 – 103,71	83,12 (88,93)	41,50 – 124,74
WUR (NRS)	1,88 (0,83)	1,49 – 2,27	2,20 (1,14)	1,66 – 2,73
PPT (kgf/cm ²)	1,79 (0,47)	1,57 – 2,01	1,80 (0,41)	1,62 – 2,01

Limite de detecção do frio (CDT), limiar de detecção do calor (WDT), limiar sensorial térmico (TSL), limiar de dor ao frio (CPT), limiar de dor ao calor (HPT), limiar de detecção mecânica (MDT), limiar de dor mecânica (MPT) razão da somação temporal (WUR) e limiar de dor a pressão (PPT). NRS = escala de classificação numérica.

Tabela 3. Análise de variância (ANOVA) considerando um fator dentro do grupo, isto é, níveis de sítios, e um fator entre grupos, isto é, sexo para diferentes parâmetros de teste sensorial quantitativo (QST).

	Efeito Principal – Sítio	Efeito Principal - Sexo	Interação – Sítio vs. Sexo
CDT °C	F=10,66, p=0,002*	F=0,92, p=0,343	F=0,70, p=0,408
WDT °C	F=14,50, p<0,001	F=0,00, p=0,968	F=0,40, p=0,532
TSL °C	F= 23,61, p<0,001	F=0,32, p=0,574	F=0,07, p=0,791
CPT °C	F=0,00, p=0,994	F=6,00, p=0,018	F= 0,00, p=0,989
HPT °C	F=8,39, p=0,006	F=8,34, p=0,006	F=0,00, p=0,951
MDT (g/mm²)	F=2,28, p=0,138	F=2,06, p=0,158	F=1,86, p=0,179
MPT (g/mm²)	F=3,53, p=0,067	F=12,8, p<0,001	F=0,07, p=0,792
WUR (NRS)	F=0,20, p=0,652	F=3,09, p=0,008	F=0,01, p=0,893
PPT (kgf/cm²)	F=1,70, p=0,199	F=15,7, p<0,001	F=5,77, p=0,021

* As células em negrito apresentam valores significativos de p (p <0,050)

Limite de detecção do frio (CDT), limiar de detecção do calor (WDT), limiar sensorial térmico (TSL), limiar de dor ao frio (CPT), limiar de dor ao calor (HPT), limiar de detecção mecânica (MDT), limiar de dor mecânica (MPT) razão da somação temporal (WUR) e limiar de dor a pressão (PPT). NRS = escala de classificação numérica.

A tabela 4 apresenta a confiabilidade entre avaliadores do QST para masseter e ATM considerando todos os participantes. Os ICCs gerais mostraram valores aceitáveis para a confiabilidade interexaminador da avaliação somatossensorial das estruturas mastigatórias. Setenta e sete por cento dos valores de ICC de todas as medidas de QST foram classificados entre razoável e excelente (razoável = 28%, bom = 16% e excelente = 33%). No entanto, a confiabilidade do WDT, TSL e HPT do masseter foi significativamente maior do que da ATM, enquanto que a confiabilidade do MDT da ATM foi maior do que do masseter (tabela 4). Os limiares de detecção térmica da ATM apresentaram erro padrão de mensuração (SEM) e menor mudança detectável (SDC) ligeiramente menores quando comparados aos valores do masseter e o oposto foi encontrado para os limiares de dor térmica, ou seja, SEM e SDC menor para o masseter (tabela 4). Para os testes mecânicos, SEM e SDC da ATM foram quase metade dos valores do masseter para MDT, enquanto SEM e SDC do masseter foram ligeiramente menor do que os valores da ATM para o MPT (tabela 4). Além disso, WUR e PPT da ATM também apresentaram SEM e SDC ligeiramente menores quando comparados aos valores do masseter (tabela 4).

Tabela 4. Confiabilidade relativa e valores de concordância da avaliação quantitativa dos testes sensoriais (QST) da articulação temporomandibular (ATM) e da região do masseter (dados masculinos e femininos combinados)

	ICC	Valor de P	ICC – 95% IC	SEM	SDC
ATM					
CDT °C	0,53	<0,001	0,27 – 0,72	1,16	3,22
WDT °C	0,58	<0,001	0,34 – 0,75	1,45	4,03
TSL °C	0,37	0,009	0,07 – 0,61	2,37	6,57
CPT °C	0,70	<0,001	0,51 – 0,83	5,45	15,11
HPT °C	0,47	0,001	0,18 – 0,68	2,97	8,24
MDT (g/mm ²)	0,97	<0,001	0,93 – 0,98	0,12	0,34
MPT (g/mm ²)	0,76	<0,001	0,59 – 0,86	32,05	88,84
WUR (NRS)	0,09	0,271	-0,19 – 0,37	0,86	2,39
PPT (kgf/cm ²)	0,60	<0,001	0,36 – 0,77	0,29	0,80
Masseter					
CDT °C	0,51	<0,001	0,24 – 0,70	1,40	3,89
WDT °C	0,83	<0,001	0,70 – 0,90	1,56	4,34
TSL °C	0,70	<0,001	0,50 – 0,83	2,65	7,33
CPT °C	0,83	<0,001	0,70 – 0,90	3,98	11,04
HPT °C	0,86	<0,001	0,75 – 0,92	1,38	3,82
MDT (g/mm ²)	0,31	0,018	0,02 – 0,56	0,23	0,63
MPT (g/mm ²)	0,82	<0,001	0,69 – 0,90	27,18	75,34
WUR (NRS)	0,08	0,272	-0,19 – 0,36	0,98	2,71
PPT (kg/cm ²)	0,56	<0,001	0,31 – 0,74	0,33	0,91

Limiar de detecção de frio (CDT), limiar de detecção ao calor (WDT), limiar sensorial térmico (TSL), limiar de dor ao frio (CPT), limiar de dor ao calor (HPT), limiar de detecção mecânica (MDT), limiar de dor mecânica (MPT) razão da somação temporal (WUR) e limiar de dor a pressão (PPT). NRS = escala de classificação numérica. ICC = coeficiente de correlação intraclasse. IC 95% = 95% de intervalo de confiança. SEM = erro padrão de mensuração. SDC = menor mudança detectável. A magnitude do ICC foi classificada como confiabilidade ruim (<0,4), razoável (0,4 - 0,59), bom (0,6 - 0,75) ou excelente (> 0,75). As IC 95% não sobrepostas foram consideradas significativamente diferentes

A Tabela 5 mostra a confiabilidade interexaminador do QST para masseter e ATM considerando homens e mulheres separadamente. Os ICCs gerais de mulheres e homens tiveram valores aceitáveis para a confiabilidade interexaminador da avaliação somatossensorial de estruturas mastigatórias, embora a distribuição dentro das categorias de magnitude do ICC tenha sido diferente (tabela 5). A seguinte combinação de teste / sítios apresentou ICCs significativamente menores para mulheres: HPT, MDT da ATM e MPT da ATM e do masseter (tabela 5). Por outro lado, o TSL da ATM apresentou ICC significativamente maior para as mulheres. Além disso, a maioria dos valores de SEM e SDC das mulheres (86%) foram apenas ligeiramente inferiores aos valores de SEM e SDC dos homens (tabela 5).

Tabela 5. Confiabilidade relativa e valores de concordância da avaliação de testes sensoriais quantitativos (QST) da articulação temporomandibular (ATM) e masseter (dados masculinos e femininos separados).

	ICC	Valor de P	ICC – 95% IC	SEM	SDC
Mulheres					
ATM					
CDT °C	0,68	<0,001	0,36 – 0,86	0,88	2,44
WDT °C	0,54	0,006	0,15 – 0,79	1,62	4,50
TSL °C	0,65	0,001	0,30 – 0,85	1,75	4,84
CPT °C	0,67	0,001	0,33 – 0,85	5,02	13,90
HPT °C	0,18	0,214	-0,28 – 0,58	3,29	9,13
MDT (g/mm ²)	0,31	0,049	-0,07 – 0,64	0,16	0,44
MPT (g/mm ²)	0,31	0,073	-0,10 – 0,64	20,52	56,88
WUR (NRS)	0,11	0,277	-0,23 – 0,47	0,89	2,48
PPT (kgf/cm ²)	0,44	0,022	0,02 – 0,73	0,25	0,70
Masseter					
CDT °C	0,65	0,001	0,30 – 0,84	1,24	3,44
WDT °C	0,86	<0,001	0,69 – 0,94	1,40	3,88
TSL °C	0,63	0,001	0,27 – 0,83	3,22	8,92
CPT °C	0,81	<0,001	0,58 – 0,92	3,94	10,93
HPT °C	0,85	<0,001	0,68 – 0,94	1,46	4,06
MDT (g/mm ²)	0,14	0,257	-0,27 – 0,53	0,22	0,61
MPT (g/mm ²)	0,31	0,090	-0,15 – 0,65	21,54	59,71
WUR (NRS)	0,14	0,233	0,22 – 0,51	0,95	2,64
PPT (kg/cm ²)	0,52	0,007	0,13 – 0,78	0,27	0,75
Homens					
ATM					
CDT °C	0,42	0,021	0,02 – 0,72	1,39	3,84
WDT °C	0,64	0,001	0,30 – 0,84	1,26	3,49
TSL °C	0,10	0,327	-0,36 – 0,52	2,86	7,93
CPT °C	0,66	0,001	0,32 – 0,85	5,86	16,23
HPT °C	0,60	0,002	0,23 – 0,82	2,61	7,25
MDT (g/mm ²)	0,99	<0,001	0,98 – 0,99	0,08	0,21
MPT (g/mm ²)	0,73	<0,001	0,44 – 0,88	40,42	112,03
WUR (NRS)	0,12	0,307	-0,35 – 0,53	0,83	2,31
PPT (kgf/cm ²)	0,45	0,019	0,03 – 0,74	0,32	0,88
Masseter					
CDT °C	0,34	0,057	-0,07 – 0,67	1,55	4,30
WDT °C	0,79	<0,001	0,53 – 0,91	1,71	4,75
TSL °C	0,81	<0,001	0,59 – 0,92	1,91	5,29
CPT °C	0,81	<0,001	0,58 – 0,92	4,02	11,15
HPT °C	0,80	<0,001	0,57 – 0,92	1,29	3,57
MDT (g/mm ²)	0,41	0,027	-0,00 – 0,71	0,23	0,65
MPT (g/mm ²)	0,84	<0,001	0,65 – 0,93	31,83	88,24
WUR (NRS)	-0,05	0,508	-0,43 – 0,42	1,00	2,78
PPT (kg/cm ²)	0,25	0,147	-0,22 – 0,62	0,38	1,05

Limiar de detecção de frio (CDT), limiar de detecção ao calor (WDT), limiar sensorial térmico (TSL), limiar de dor ao frio (CPT), limiar de dor ao calor (HPT), limiar de detecção mecânica (MDT), limiar de dor mecânica (MPT) razão da somação temporal (WUR) e limiar de dor a pressão (PPT). NRS = escala de classificação numérica. ICC = coeficiente de correlação

intraclasse. IC 95% = 95% de intervalo de confiança. SEM = erro padrão de medição. SDC = menor alteração detectável. A magnitude do ICC foi classificada como confiabilidade ruim ($<0,4$), razoável (0,4 - 0,59), bom (0,6 - 0,75) ou excelente ($>0,75$). As IC 95% não sobrepostas foram consideradas significativamente diferentes

6 Discussão

O presente estudo estimou a confiabilidade relativa interexaminador e a concordância da avaliação de QST das estruturas mastigatórias, e os principais achados foram: a) apesar dos limiares térmicos relacionados ao calor (WDT, TSL e HPT) e detecção tátil (MDT) demonstrarem diferenças significativas nos valores de ICC em relação ao sítio de avaliação, os valores gerais da confiabilidade relativa interexaminador da avaliação somatossensorial do QST no masseter e na ATM são aceitáveis para pesquisas e aplicações clínicas; b) a confiabilidade relativa dos limiares mecânicos (MDT e MPT) foi significativamente menor para mulheres; c) os níveis mais baixos de concordância foram apresentados para o CPT e MPT.

A avaliação somatossensorial, independentemente do quão padronizada e sofisticada, pode ser influenciada por vários fatores: atenção, cognição, motivação, experiência e habilidade do examinador, ansiedade e estresse, podendo ser impraticável controlar antecipadamente todos eles (Backonja et al., 2013; Backonja et al., 2009). No entanto, evidências recentes tentaram estimar os componentes de variância considerando múltiplas fontes de variabilidade e mostraram que a principal variação está relacionada às diferenças reais entre os participantes (Moana-Filho et al., 2017; O'Neill e O'Neill 2015). Assim, nossos achados gerais estão em acordo com os estudos anteriores de que uma avaliação somatossensorial orofacial reprodutível pode ser alcançada por meio da bateria padronizada do DFNS (Baad-Hansen et al., 2015; Geber et al., 2011; Marcuzzi et al., 2017; Pigg et al., 2010; Svensson, 2017). Uma exceção importante é o WUR, que em nosso caso foi realizada com um conjunto padronizado de monofilamentos Semmes-Weinstein. No entanto, uma evidência anterior que avaliou o WUR por meio de pontas metálicas padronizadas também relatou fraca confiabilidade relativa, o que indica que a somação temporal avaliada pelo protocolo DFNS pode representar um fenômeno instável (Geber et al., 2011).

Apesar das evidências acima mencionadas em favor da aplicabilidade clínica do QST na região orofacial, nossos resultados também chamam a atenção para alguns detalhes importantes dentro do contexto geral de reprodutibilidade da avaliação somatossensorial. Primeiro, a heterogeneidade entre os sítios pode afetar significativamente a confiabilidade relativa, o que significa que as respostas do QST de diferentes sítios não são comparáveis em sua capacidade discriminatória. Embora uma sensibilidade específica entre regiões tenha sido mostrada quando comparado áreas de regiões espinhais e áreas trigeminais (Rolke et al., 2006),

nossos resultados indicam que mesmo dentro do território trigeminal as especificidades dos sítios devem ser levadas em consideração. Por exemplo, apesar da variabilidade entre dois examinadores, os testes térmicos realizados no masseter possuem uma melhor capacidade de discriminação entre sujeitos em comparação com os testes térmicos aplicados na ATM. Considerando a concordância geral dos testes térmicos entre masseter e ATM, essas diferenças de confiabilidade podem ser devido a uma menor variabilidade de valores dos testes sensoriais da ATM entre os participantes, o que pode ser constatado a partir de um maior desvio padrão nas medições do masseter quando comparado com os da ATM (de Vet et al., 2006). Uma possível explicação para essas diferenças pode ser a menor e mais bem definida área anatômica da ATM, que pode apresentar limiares menos variáveis. Assim, testes térmicos na ATM podem ser mais propensos à erros de mensuração associados aos diferentes examinadores. Por outro lado, a baixa variabilidade entre os participantes e o maior erro de mensuração estão relacionados à fraca confiabilidade relativa da detecção tátil no masseter. No entanto, esta fraca confiabilidade interexaminador do masseter é contrastada com a boa confiabilidade teste-reteste relatada em uma investigação anterior do nosso grupo (Costa et al., 2017).

A confiabilidade relativa depende da heterogeneidade da população da qual a amostra é originária (de Vet et al., 2006). Isso significa que a confiabilidade de uma amostra menos heterogênea é mais sensível aos erros de mensuração, ou seja, quando a variação entre os indivíduos é baixa, um pequeno erro de medição dificulta a distinção entre os sujeitos (de Vet et al., 2006). No entanto, a maioria das publicações sobre a confiabilidade do QST considerou as diferenças de heterogeneidade apenas entre pacientes com condições dolorosas de interesse e / ou participantes saudáveis (Andersen et al., 2015; Baad-Hansen et al., 2015; Geber et al., 2011; Marcuzzi et al., 2017; Pigg et al., 2010). Portanto, até onde sabemos, este é o primeiro estudo que apresenta evidências de que as diferenças sexuais na sensibilidade somatossensorial podem influenciar a confiabilidade dos limiares mecânicos. As participantes femininas apresentaram valores de ICC significativamente menores para MDT e MPT quando comparados aos homens. Do mesmo modo, muitas das diferenças de confiabilidade relativa provavelmente estão relacionadas à menor variabilidade entre as respostas do QST das mulheres, o que pode ser visto a partir do maior desvio padrão para os valores de MDT e MPT da

amostra masculina quando comparado às mulheres. Os valores de referência do DFNS também apresentaram um desvio padrão maior para MDT e MPT em um grupo de jovens adultos do sexo masculino (menos de 40 anos) (Rolke et al., 2006). Além disso, um estudo anterior das diferenças sexuais na sensibilidade somatossensorial das zonas de entrada do nervo trigeminal do nervo infraorbitário e do nervo alveolar inferior mostrou uma maior variabilidade do MPT no grupo masculino (Matos et al., 2011). A vasta literatura sobre diferenças de sexo na dor tem mostrado que as mulheres podem ser mais sensíveis a algumas modalidades de estímulos de dor evocada do que os homens e uma complexa interação biológica, por exemplo, capacidade moduladora endógena da dor, hormônios sexuais, fatores sociais e culturais podem explicar essas diferenças (Fillingim 2017; Greenspan et al., 2007; Racine et al., 2012; Riley et al., 1998). Tais diferenças podem levar a este aumento da variabilidade das respostas do QST, que podem afetar de forma significativa a capacidade de discriminação em testes que implicam uma maior interação entre o examinador e o participante, o que pode ocorrer nos testes mecânicos. No entanto, é necessária uma investigação mais aprofundada para corroborar e elucidar essas descobertas e seus possíveis mecanismos subjacentes.

Os parâmetros de concordância são considerados uma medição negligenciada em estudos de propriedades métricas de instrumentos médicos (de Vet et al., 2006). Em particular, uma quantidade considerável de estudos de confiabilidade do QST se concentra principalmente em parâmetros de confiabilidade relativa, por exemplo, coeficientes de correlação (Baad-Hansen et al., 2015; Costa et al., 2017; Geber et al., 2011; Pigg et al 2010). Nossas descobertas gerais defendem a aplicabilidade clínica do QST considerando seus níveis aceitáveis de confiabilidade relativa. No entanto, uma avaliação abrangente da capacidade de reprodutibilidade do QST deve incluir parâmetros de concordância. Nesse sentido, nossos resultados mostraram uma concordância geral ligeiramente maior para a ATM quando comparado aos valores do masseter e das mulheres quando comparadas aos homens. No entanto, o CPT e o MPT, independentemente do sítio e do sexo, tiveram os níveis menores de concordância, o que é esperado, considerando que a concordância está mais relacionada ao próprio instrumento e é menos propensa à heterogeneidade de diferentes populações / sítios. Por exemplo, uma investigação recente da confiabilidade teste-reteste a longo prazo do QST

também encontrou níveis baixos de concordância para o CPT aplicado na mão de participantes saudáveis (Marcuzzi et al., 2017). Do mesmo modo, CPT e MPT apresentaram alta variabilidade em pacientes com dor no ombro (Lindgren et al., 2016). Tais níveis de concordância de CPT e MPT significam que as medições repetidas realizadas por diferentes examinadores são menos precisas para fins de avaliação e monitoramento. Além disso, nossos valores de concordância fornecem informações úteis sobre as mudanças clínicas relevantes que devem ser esperadas quando os resultados de diferentes examinadores são levados em consideração.

Os pontos fortes desta investigação foram a aplicação de abordagens estatísticas bem estabelecidas para estimar a reprodutibilidade, ou seja, ICC e SEM na avaliação do QST em estruturas mastigatórias musculares e articulares. No entanto, algumas limitações também precisam ser destacadas: a) a amostra foi composta por jovens adultos, que pode dificultar a validade externa para populações mais velhas, como adultos com mais de 40 anos; b) pacientes com DTM não foram avaliados, o que também é importante, considerando que a confiabilidade relativa depende das características da população; c) a bateria completa de QST não foi possível ser realizada devido a falta de alguns equipamentos em nosso laboratório; d) não foi realizada a confiabilidade teste-reteste, que também é uma fonte importante de erros de mensuração.

7 Conclusões

7 CONCLUSÕES

A avaliação somatossensorial geral das estruturas mastigatórias realizadas por dois examinadores pode ser considerada suficientemente confiável para discriminar os participantes. No entanto, influências na reprodutibilidade dos parâmetros somatossensoriais, originadas por diferenças no sítio de avaliação e no sexo dos participantes, devem ser levadas em consideração para uma interpretação e aplicação clínica apropriadas.

Referências

REFERÊNCIAS

Andersen, K. G., Kehlet, H., & Aasvang, E. K. (2015). Test-retest agreement and reliability of quantitative sensory testing 1 year after breast cancer surgery. *The Clinical journal of pain*, 31(5), 393-403.

Arendt-Nielsen, L., & Yarnitsky, D. (2009). Experimental and clinical applications of quantitative sensory testing applied to skin, muscles and viscera. *The Journal of Pain*, 10(6), 556-572.

Baad-Hansen, L., Arima, T., Arendt-Nielsen, L., Neumann-Jensen, B., & Svensson, P. (2010). Quantitative sensory tests before and 1½ years after orthognathic surgery: a cross-sectional study. *Journal of oral rehabilitation*, 37(5), 313-321.

Baad-Hansen, L., Pigg, M., Yang, G., List, T., Svensson, P., & Drangsholt, M. (2015). Reliability of intra-oral quantitative sensory testing (QST) in patients with atypical odontalgia and healthy controls—a multicentre study. *Journal of oral rehabilitation*, 42(2), 127-135.

Backonja, M. M., Attal, N., Baron, R., Bouhassira, D., Drangholt, M., Dyck, P. J. & Hansson, P. (2013). Value of quantitative sensory testing in neurological and pain disorders: NeuPSIG consensus. *Pain*, 154(9), 1807-1819.

Backonja, M. M., Walk, D., Edwards, R. R., Sehgal, N., Moeller-Bertram, T., Wasan, A., ... & Wallace, M. (2009). Quantitative sensory testing in measurement of neuropathic pain phenomena and other sensory abnormalities. *The Clinical journal of pain*, 25(7), 641-647.

Bair, E., Gaynor, S., Slade, G. D., Ohrbach, R., Fillingim, R. B., Greenspan, J. D. & Maixner, W. (2016). Identification of clusters of individuals relevant to temporomandibular disorders and other chronic pain conditions: the OPPERA study. *Pain*, 157(6), 1266.

Bannigan, K., & Watson, R. (2009). Reliability and validity in a nutshell. *Journal of clinical nursing*, 18(23), 3237-3243.

Baron, R., Maier, C., Attal, N., Binder, A., Bouhassira, D., Cruccu, G. & Jensen, T. S. (2017). Peripheral neuropathic pain: a mechanism-related organizing principle based on sensory profiles. *Pain*, 158(2), 261.

Costa, Y. M., Morita-Neto, O., Araújo-Júnior, E. N. S., Sampaio, F. A., Conti, P. C. R., & Bonjardim, L. R. (2017). Test–retest reliability of quantitative sensory testing for mechanical somatosensory and pain modulation assessment of masticatory structures. *Journal of oral rehabilitation*, 44(3), 197-204.

De Leeuw, R. (2010). *Dor orofacial: guia de avaliação, diagnóstico e tratamento*. São Paulo: Quintessence, 315.

de Vet, H. C., Terwee, C. B., Knol, D. L., & Bouter, L. M. (2006). When to use agreement versus reliability measures. *Journal of clinical epidemiology*, 59(10), 1033-1039.

Fillingim, R. B. (2017). Individual differences in pain: understanding the mosaic that makes pain personal. *Pain*, 158, S11-S18.

Geber, C., Klein, T., Azad, S., Birklein, F., Gierthmühlen, J., Hüge, V. & Baron, R. (2011). Test–retest and interobserver reliability of quantitative sensory testing according to the protocol of the German Research Network on Neuropathic Pain (DFNS): a multi-centre study. *PAIN*, 152(3), 548-556.

Greenspan, J. D., Craft, R. M., LeResche, L., Arendt-Nielsen, L., Berkley, K. J., Fillingim, R. B., ... & Mogil, J. S. (2007). Studying sex and gender differences in pain and analgesia: a consensus report. *Pain*, 132, S26-S45.

Greenspan, J. D., Slade, G. D., Bair, E., Dubner, R., Fillingim, R. B., Ohrbach, R. & Maixner, W. (2013). Pain sensitivity and autonomic factors associated with development of TMD: the OPPERA prospective cohort study. *The Journal of Pain*, 14(12), T63-T74.

Hilgenberg-Sydney, P. B., Kowacs, P. A., & Conti, P. C. R. (2016). Somatosensory evaluation in dysfunctional syndrome patients. *Journal of oral rehabilitation*, 43(2), 89-95.

Kothari, S. F., Baad-Hansen, L., Oono, Y., & Svensson, P. (2015). Somatosensory assessment and conditioned pain modulation in temporomandibular disorders pain patients. *Pain*, 156(12), 2545-2555.

Lindgren, I., Ekstrand, E., & Brogårdh, C. (2016). Measurement variability of quantitative sensory testing in persons with post-stroke shoulder pain. *Journal of rehabilitation medicine*, 48(5), 435-441.

Maier, C., Baron, R., Tölle, T. R., Binder, A., Birbaumer, N., Birklein, F., & Krumova, E. K. (2010). Quantitative sensory testing in the German Research Network on Neuropathic Pain (DFNS): somatosensory abnormalities in 1236 patients with different neuropathic pain syndromes. *Pain*, 150(3), 439-450.

Marcuzzi, A., Wrigley, P. J., Dean, C. M., Adams, R., & Hush, J. M. (2017). The long-term reliability of static and dynamic quantitative sensory testing in healthy individuals. *Pain*, 158(7), 1217-1223.

Matos, R., Wang, K., Jensen, J. D., Jensen, T., Neuman, B., Svensson, P., & Arendt-Nielsen, L. (2011). Quantitative sensory testing in the trigeminal region: site and gender differences. *Journal of orofacial pain*, 25(2).

Moana-Filho EJ, Alonso AA, Kapos FP, Leon-Salazar V, Durand SH, Hodges JS, Nixdorf DR. Multifactorial assessment of measurement errors affecting intraoral quantitative sensory testing reliability. *Scand J Pain* 2017;16:93-98.

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. & de Vet, H. C. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of clinical epidemiology*, 63(7), 737-745.

Moloney, N. A., Hall, T. M., & Doody, C. M. (2012). Reliability of thermal quantitative sensory testing: a systematic review. *Journal of rehabilitation research and development*, 49(2), 191.

Nothnagel, H., Puta, C., Lehmann, T., Baumbach, P., Menard, M. B., Gabriel, B. & Musial, F. (2017). How stable are quantitative sensory testing measurements over time? Report on 10-week reliability and agreement of results in healthy volunteers. *Journal of pain research*, 10, 2067.

O'Neill, S., & O'Neill, L. (2015). Improving QST reliability—more raters, tests, or occasions? A multivariate generalizability study. *The Journal of Pain*, 16(5), 454-462.

Oono, Y., Wang, K., Baad-Hansen, L., Futarmal, S., Kohase, H., Svensson, P., & Arendt-Nielsen, L. (2014). Conditioned pain modulation in temporomandibular disorders (TMD) pain patients. *Experimental brain research*, 232(10), 3111-3119.

Pfau, D. B., Rolke, R., Nickel, R., Treede, R. D., & Daublaender, M. (2009). Somatosensory profiles in subgroups of patients with myogenic temporomandibular disorders and fibromyalgia syndrome. *PAIN*, 147(1-3), 72-83.

Pigg, M., Baad-Hansen, L., Svensson, P., Drangsholt, M., & List, T. (2010). Reliability of intraoral quantitative sensory testing (QST). *Pain*, 148(2), 220-226.

Racine, M., Tousignant-Laflamme, Y., Kloda, L. A., Dion, D., Dupuis, G., & Choinière, M. (2012). A systematic literature review of 10 years of research on sex/gender and experimental pain perception—Part 1: Are there really differences between women and men?. *Pain*, 153(3), 602-618.

Riley III, J. L., Robinson, M. E., Wise, E. A., Myers, C. D., & Fillingim, R. B. (1998). Sex differences in the perception of noxious experimental stimuli: a meta-analysis. *Pain*, 74(2-3), 181-187.

Rolke, R., Baron, R., Maier, C. A., Tölle, T. R., Treede, R. D., Beyer, A., ... & Braune, S. (2006). Quantitative sensory testing in the German Research Network on Neuropathic Pain (DFNS): standardized protocol and reference values. *Pain*, 123(3), 231-243.

Schiffman, E., Ohrbach, R., Truelove, E., Look, J., Anderson, G., Goulet, J. P., ... & Svensson, P. (2014). 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 Group. *Journal of oral & facial pain and headache*, 28(1), 6.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin*, 86(2), 420.

Stratford, P. W., & Goldsmith, C. H. (1997). Use of the standard error as a reliability index of interest: an applied example using elbow flexor strength data. *Physical therapy*, 77(7), 745-750.

Svensson, P. (2017). Intraoral QST—Mission impossible or not?. *Scandinavian journal of pain*, 16(1), 112-113.

Svensson, P., BAAD-HANSEN, L., Pigg, M., List, T., Eliav, E., Ettlin, D., ... & Essick, G. (2011). Guidelines and recommendations for assessment of somatosensory function in oro-facial pain conditions—a taskforce report. *Journal of oral rehabilitation*, 38(5), 366-394.

Uddin, Z., & MacDermid, J. C. (2016). Quantitative sensory testing in chronic musculoskeletal pain. *Pain Medicine*, 17(9), 1694-1703.

von Hehn CA, Baron R, Woolf CJ. Deconstructing the neuropathic pain phenotype to reveal neural mechanisms. *Neuron* 2012;73(4):638-652.

Walk, D., Sehgal, N., Moeller-Bertram, T., Edwards, R. R., Wasan, A., Wallace, M. & Backonja, M. M. (2009). Quantitative sensory testing and mapping: a review of nonautomated quantitative methods for examination of the patient with neuropathic pain. *The Clinical journal of pain*, 25(7), 632-640.

Yang, G., Baad-Hansen, L., Wang, K., Fu, K., Xie, Q. F., & Svensson, P. (2016). Somatosensory abnormalities in Chinese patients with painful temporomandibular disorders. *The journal of headache and pain*, 17(1), 31.