95 AÑOS Facultad de Odontología

Odontoestomatología
DOI: 10.22592/ode2025n45e338

Comparison of two diagnostic tests for the detection of temporomandibular disorders. A systematic review and meta-analysis.

UPDATE

Comparación de dos pruebas diagnósticas para la detección de trastornos temporomandibulares. Una revisión sistemática y metanálisis.

Comparação de dois testes de diagnóstico para a detecção de distúrbios temporomandibulares. Uma revisão sistemática e meta-análise

Abstract

The aim of this review was to assess the validity of two screening instruments for temporomandibular disorders (the FAI and the 3Q-TMD test) that have had the RDC and DC TMD tests as reference examinations.

Six electronic databases were searched from 1992, the date of publication of the RDC-TMD, to 14 April 2022. Two independent reviewers selected the studies. Risk of bias and applicability were assessed using the QUADAS 2 instrument. Out of 798 articles, 10 were included representing a total of 4106 subjects. Five studies were considered at high risk of bias, four at low risk and one unclear. The meta-analysis showed a sensitivity of 0.92 and a specificity of 0.79.

As a limitation, a large methodological heterogeneity and sample characteristics were detected, however, it is possible to conclude that the FAI and 3Q-TMD instruments are valid for detection of individuals with TMD.

Hugo Daniel Rivera[1]
Claudia Ivonne Rodríguez[2]
Ramón Álvarez-Vaz[3]
Marcelo Kreiner[4]

CORRESPONDENCE
Hugo Daniel Rivera:
hugo.rivera@uap.edu.ar

Received: April 10, 2024
Accepted: November 05, 2024

Key words: Craniomandibular Disorders. Temporomandibular joint disorders. Triage. Orofacial pain.

1 Cátedra de Oclusión.Facultad de Ciencias de la Salud. Universidad Adventista del Plata.

2 Laboratorio de Fisiología. Facultad de Odontología Universidad Nacional Autónoma de México.

3 Doctor en Cs. Médicas, Prof. Agdo, Facultad de Odontología, y Facultad de Ciencias Económicas y de Administración, Universidad de la República.

4 Cátedra de Fisiología. Facultad de Odontología. Universidad de la República Uruguay.

Resumen

El objetivo de esta revisión fue evaluar la validez de dos instrumentos de detección de trastornos temporomandibulares, el IAF y el test 3Q-TMD, que hayan tenido como exámenes de referencia las pruebas RDC y DC TMD.

Se realizaron búsquedas desde el año 1992 fecha en que se publican los RDC-TMD hasta el 14 de abril de 2022 en 6 bases de datos electrónicas. Dos revisores independientes seleccionaron los estudios. El riesgo de sesgo y aplicabilidad se evaluó mediante el instrumento QUADAS 2. De 798 artículos, se incluyeron 10 que representaron un total de 4106 sujetos. Cinco estudios se consideraron de alto riesgo de sesgo, cuatro de bajo riesgo y uno no claro. El metaanálisis mostró una sensibilidad de 0.92 y una especificidad de 0.79.

Como limitación se detectó una gran heterogeneidad metodológica y de características de la muestra, sin embargo, es posible concluir que los instrumentos IAF y 3Q-TMD son válidos para detección de individuos con TTM.

Palabras claves: Desórdenes Craniomandibulares, Trastornos Craneomandibulares, Trastornos de la articulación Temporomandibular, triaje, dolor orofacial.

Resumo

O objetivo desta revisão foi avaliar a validade de dois instrumentos de triagem para distúrbios temporomandibulares (o IAF e o teste 3Q-TMD) que tiveram os testes RDC e DC TMD como exames de referência.

Seis bancos de dados eletrônicos foram pesquisados de 1992, a data de publicação do RDC-TMD, até 14 de abril de 2022. Dois revisores independentes selecionaram os estudos. O risco de viés e a aplicabilidade foram avaliados usando o instrumento QUADAS 2. Dos 798 artigos, 10 foram incluídos, representando um total de 4.106 indivíduos. Cinco estudos foram considerados de alto risco de viés, quatro de baixo risco e um não claro. A meta-análise mostrou uma sensibilidade de 0.92 e uma especificidade de 0.79.

Como limitação, foi detectada uma grande heterogeneidade metodológica e características da amostra; no entanto, é possível concluir que os instrumentos IAF e 3Q-TMD são válidos para a detecção de indivíduos com DTM.

Palavras-chave: Distúrbios Craniomandibulares. Disfunções Craniomandibulares. Transtornos da Articulação Temporomandibular. Triagem. Dor Orofacial.

Introdution

Temporomandibular disorders (TMDs) are conditions clinically characterized by pain and/or dysfunction in the masticatory, cervical, and head muscles, temporomandibular joints (TMJs), and adjacent structures [1]. TMDs represent a significant public health concern, affecting approximately 30% of the general population [2], and are considered the most common cause of chronic non-dental pain in the orofacial region [3]. TMD-related pain can impair an individual's ability to perform daily activities, as well as their psychosocial functioning and quality of life[4]. However, despite their negative impact on daily life, these conditions often go undetected or overlooked in routine dental care, as evidenced by the gap between the estimated need for treatment and the actual treatments performed [5,6]. Patients typically seek treatment only when the pain becomes very intense or chronic [7,8].Studies on the economic and occupational implications of chronic painful conditions, such as headaches or facial pain caused by TMDs, highlight their significant burden. Patients with TMDs are estimated to incur 50% higher average costs for medication and professional consultations [9]. In the United States alone, the cost of treating these conditions has reportedly doubled over the past decade, reaching approximately $4 billion per year [10]. Various studies indicate that early intervention in these disorders achieves high success rates and reduces long-term treatment costs [11–13].

Incorporating routine TMD screening tests into dental practice would facilitate timely diagnosis and treatment for these patients, thereby improving their quality of life and reducing treatment expenses.

TMD DIAGNOSTIC TESTS

There are several diagnostic systems for orofacial pain caused by TMD [14,15]. The Diagnostic Criteria for TMD (DC/TMD) and its earlier version, the RDC/TMD, are strictly defined diagnostic methods for the most common TMD conditions, such as myalgia, arthralgia, myofascial pain, degenerative joint disorders, and functional disorders of the temporomandibular joint. Their evaluation system, which separates psychosocial aspects from phys-

ical aspects, has been considered a paradigm shift within the field of orofacial pain [16].

In recent years, the RDC/TMD and DC/TMD diagnostic protocols have been used as reference tests or "gold standards" to evaluate screening tests for the detection of TMD. Although DC/TMD is reliable and valid, its routine use in clinical TMD triage is not practical, as its evaluation protocol is time-consuming and requires proper interpretation of complex algorithms [17]. In this sense, TMD screening tools should be cost-effective, simple, efficient, and accurate.

SCREENING TESTS FOR DETECTION OF TMD

Early detection has been suggested to prevent chronic conditions [13]. Additionally, most patients diagnosed with TMD would benefit from conservative and single-treatment approaches [18].

Several diagnostic methods have been validated for the detection of TMD. Among the most studied are the TMD pain screening from the RDC/TMD test, the Fonseca Anamnestic Index (FAI) [19], and the 3Q-TMD test [20]. Of these three tests, RDC/TMD screening has proven valid for identifying individuals with potential TMD pain; however, it does not account for non-painful functional aspects [16].

The FAI is a questionnaire consisting of 10 items in its original version and 5 items in its short form (SFAI), both designed to detect TMD. The items are scored to reflect the severity of the disorder. The FAI has shown consistent results with other tools for detecting TMD, including the questionnaire from the American Academy of Orofacial Pain, and has been evaluated against RDC/TMD and DC/TMD [21,22].

In contrast, the 3Q-TMD test is shorter, comprising a 3-item questionnaire. Two items focus on facial pain, while the third addresses jaw locking during function. Given that patients with TMD often go undetected and untreated in dental practice, this study focuses on two screening methods that could aid in identifying the most common TMDs, whether painful or functional. Therefore, this systematic review aims to compare these two instruments by addressing the following question: What is the validity (specificity and sensitivity: accuracy) of the FAI and 3Q-TMD instruments, as used in clinical and epidemiological settings, for detecting TMDs?

Methods

PROTOCOL AND REGISTRATION

The protocol for this review was developed following the PRISMA-P reporting guideline [23] and was submitted for registration with the International Prospective Register of Systematic Reviews (PROSPERO, Centre for Reviews and Dissemination, University of York; and the National Institute for Health Research). To conduct this systematic review of diagnostic tests, the PRISMA-DTA checklist recommendations were followed [24].

ELIGIBILITY CRITERIA

This review included diagnostic accuracy studies for the detection of TMD based on the FAI as an index test and the 3Q-TMD test as a comparator. For both tests, studies using the RDC/TMD and DC/TMD as the gold standard were examined. No restrictions were applied regarding the age or sex of participants or the publication language of the studies. Both "cohort" and "case-control" study designs were accepted.

As regards sample collection sites, studies conducted in both primary care (general dental centers) and secondary care (specialized orofacial pain centers) settings were included. To establish the target condition, patients in these studies underwent an individual clinical assessment of TMD (specifically joint noises, limitations in jaw movement, muscle pain, joint pain, and/or preauricular pain) following the criteria established by the INFORM consortium [25].

Exclusion criteria were as follows:

- Studies that did not use the RDC/TMD (i.e., studies published before 1992) or DC/TMD, or studies that modified the tool;
- Articles containing duplicate data from another included study;
- Studies that were not exclusively focused on diagnosing patients with TMD;
- Reviews, letters, books, expert opinions, and case reports.

IDENTIFICATION OF STUDY SOURCES

The HIRU search filter for diagnostic accuracy studies was used [26]. An electronic search strategy was subsequently developed for PubMed and adapted for each of the following bibliographic databases: Web of Science, ScienceDirect, Scopus, and Scielo. Grey literature databases such as Google Scholar and ProQuest One Academic were also searched.

The search period covered the years from 1992 to April 2022. Additionally, the reference lists of included studies were manually reviewed to identify any additional relevant studies. A reference manager* was used to compile references and remove duplicates (*Mendeley®, Elsevier Amsterdam, The Netherlands).

SEARCH STRATEGY

Indexed terms and free terms were used to locate research conducted on TMD diagnostic studies, as well as their diagnostic accuracy. Below is a description of the search strategy using the filter for diagnostic accuracy studies [26]:

Search: ((“temporomandibular joint disorders”[mh]
OR “Disorder, Temporomandibular Joint”[tiab]
OR “Disorders, Temporomandibular Joint”[tiab]
OR “Joint Disorders, Temporomandibular”[tiab]
OR “Temporomandibular Joint Disorder”[tiab]
OR “TMJ Disorders”[tiab]
OR “Disorder, TMJ”[tiab]
OR “Disorders, TMJ”[tiab]
OR “TMJ Disorder”[tiab]
OR “Temporomandibular Disorders”[tiab]
OR “Disorder, Temporomandibular”[tiab]
OR “Disorders, Temporomandibular”[tiab]
OR “Temporomandibular Disorder”[tiab]
OR “Temporomandibular Joint Diseases”[tiab]
OR “Disease, Temporomandibular Joint”[tiab]
OR “Diseases, Temporomandibular Joint”[tiab]
OR “Temporomandibular Joint Disease”[tiab]
OR “TMJ Diseases”[tiab]
OR “Disease, TMJ”[tiab]
OR “Diseases, TMJ”[tiab]
OR “TMJ Disease”[tiab]) AND (Diagnosis[mh]
OR diagnosis[tiab]
OR Triage[tiab])) AND (sensitiv*[TiAb]
OR sensitivity and specificity[mh]
OR (predictive[Tiab] AND value*[Tiab])
OR predictive value of tests[mh]
OR accuracy*[Tiab]).

STUDY SELECTION

In phase 1, two authors (HDR and CIR) independently evaluated the titles and abstracts of the identified studies by applying the previously established eligibility criteria. Once the articles were deemed eligible for inclusion, the reviewers performed a full-text screening in phase 2. Finally, articles that did not meet the eligibility criteria were removed. Although a third reviewer (MK) was available to intervene in the event of potential disagreements, these were settled among the authors in several consensus meetings.

DATA COLLECTION PROCESS

Data collection was performed by the first reviewer (HDR) and verified by the second reviewer (CIR) to ensure the integrity of the contents. Disagreements were settled by consensus with the third reviewer (MK). A spreadsheet was used, and the following data were extracted for each included study:

Bibliometric data: year of publication, country of origin, author, publication journal.

Methodological data: total number of participants in each study, age, sex, patient blinding, and randomization method used. Study health care setting (primary or secondary care).

Statistical data: the following data were extracted: sample size of each group, sensitivity and specificity data, predictive values, diagnostic OR, and area under the ROC curve.

RISK OF BIAS

The risk of bias was independently assessed by two reviewers (HDR and CIR). The QUADAS 2 tool was used to evaluate the risk of bias and applicability of the studies [27]. The evaluation covered four distinct domains: patient selection, index test, reference standard, and patient flow and timing throughout the study. Before applying the tool, a pilot test was conducted to ensure consensus on the assessment of risk of bias between the two reviewers. No overall summary score was calculated; however, for each domain, concerns regarding bias and applicability were rated as “low,” “high,” or “unclear.” Discrepancies were resolved through consensus. These results are fully detailed in an appended paper.

DIAGNOSTIC ACCURACY MEASURES

Data from 2x2 tables were used to calculate sensitivity, specificity, and predictive values for each study. The results of individual studies were graphically presented by plotting sensitivity and specificity estimates (along with their 95% confidence intervals). Additional estimates, such as the diagnostic Odds Ratio, were calculated, and a proportional hazards model was used to estimate the performance of the tests on a receiving operating characteristic (ROC) curve. This approach provided a single value representing the overall detection ability of the screening tests.

RESULTS SYNTHESIS

The results were summarized by plotting estimates of sensitivity and specificity for the index and comparator tests (FAI and 3Q-TMD) on coupled forest plots and a matrix of receiver operating characteristics (ROC curves).

META-ANALYSIS

Summary operating points (summarized sensitivities and specificities) were estimated for both tests with a 95% confidence interval. Subsequently, the diagnostic odds ratio was determined to express the diagnostic accuracy of each test as a single number. Finally, the area under the curve for both tests was calculated on a receiver operating characteristic (ROC) curve using a proportional hazards model.

Results

STUDY SELECTION

A total of 798 articles were identified. Of these, 727 articles were excluded due to ineligible titles that did not align with the search objectives. At this stage, the main reason for excluding titles was that the studies did not pertain to screening tests. After selecting 71 studies for title and abstract evaluation, 49 duplicate studies were removed (phase 1). Ultimately, 22 studies were deemed eligible for full evaluation. After the full-text screening (phase 2), 12 studies were excluded because they did not focus on the evaluation of screening tests. Finally, 10 articles were included in the systematic review and meta-analysis. An overview of the selection process is presented in **Figure 1**.

CHARACTERISTICS OF THE STUDIES

The 10 included studies [17, 19, 20, 28–34] involved a total of 4,106 subjects (2,348 females, 580 males, 878 with unspecified gender), with a male-to-female ratio of 1 to 4. The age range of participants in these studies spanned from 11 to 78 years, with a mean age of 3.7 years.

The studies were conducted across 7 different countries, with sample sizes ranging from 102 to 923 participants. The descriptive characteristics of these studies are summarized in **Table 1**.

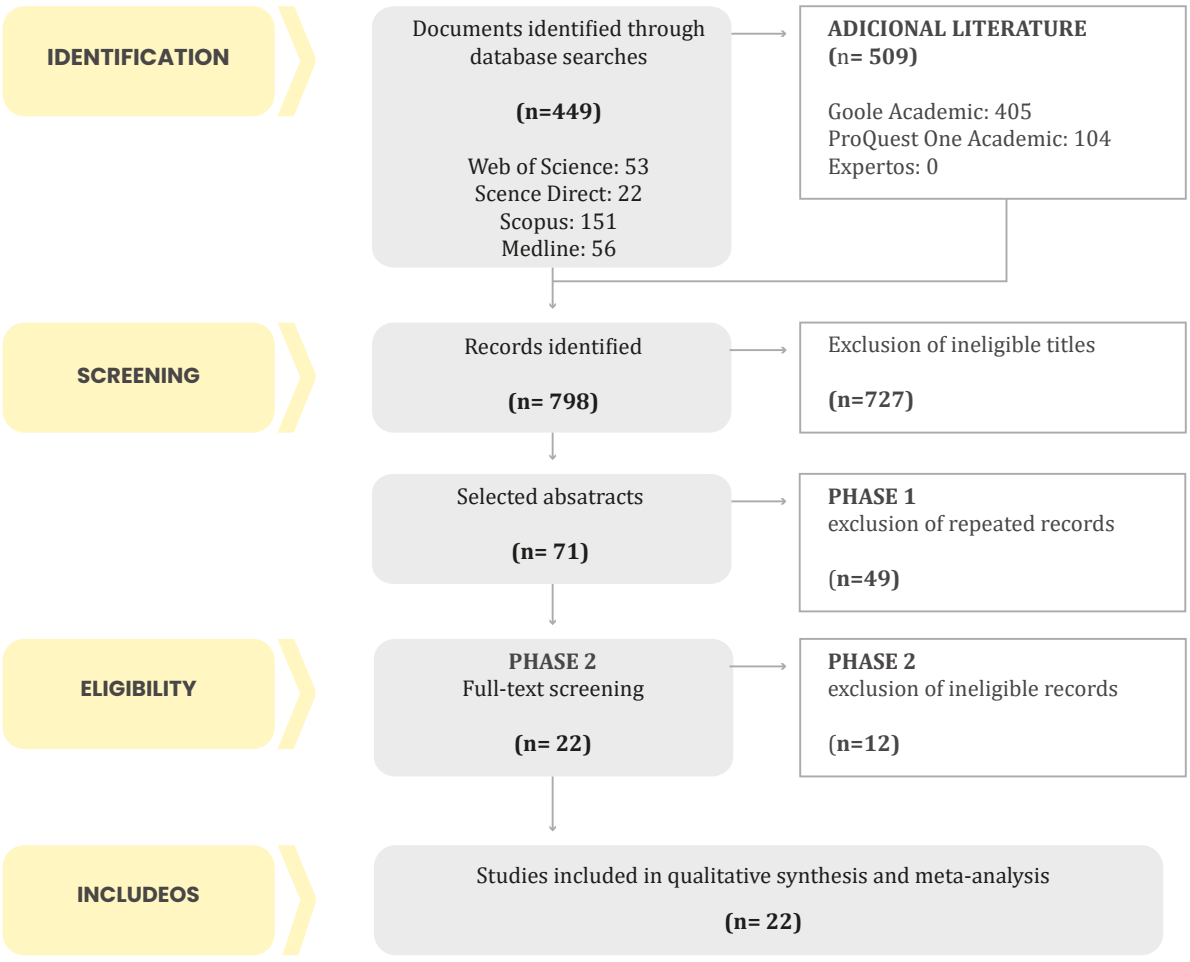

Figure 1 Flow diagram of the selection process

TABLE 1
Summary of the Main Characteristics of the Studies

AUTHORS/ LOCATION	Nº OF PARTICIPANTS		SAMPLING TYPE	SETTING	STUDY DESIGN	INDEX TEST / COMPARATOR	REFERENCE TEST
	Men	Women					
Lovgren et al. Suecia	126	323	Consecutive	Secondary (Orofacial pain center)	Observational Case-Control	3Q-TMD Comparator	DC-TMD
Lovgren et al. Suecia	300		Convenience	Primary	Observational Case-Control	3Q-TMD Comparator	DC-TMD
Stasiak et al. Brasil	265		Convenience	Secondary	Observational Cohort	FAI Index	RDC-TMD
Kaynak et al. Turquia	45	160	Convenience	Primary	Observational Case-Control	FAI Index	RDC-TMD
Zhang et al. China	613		Convenience	Primary	Observational Case-Control	FAI Index	DC-TMD
Dos Santos et al. Brasil	0	203	Convenience	Primary	Observational Case-Control	FAI Index	DC-TMD
Ujin et al. China	192	731	Consecutive	Secondary	Observational Case-Control	FAI Index	DC-TMD
Zagalaz - Anula et al. España	25	77	Convenience	Secondary	Observational Case-Control	SFAI Index	DC-TMD
Ujin et al. China	192	731	Consecutive	Secondary	Observational Case-Control	SFAI Index	DC-TMD
Pire et al. Brasil	0	123	Consecutive	Secondary	Observational Case-Control	SFAI Index	RDC-TMD

RISK OF BIAS AND APPLICABILITY OF THE STUDIES

Due to the methodological differences among the included studies, it was necessary to assess the factors determining their internal and external validity. The QUADAS 2 tool [27] allowed to determine the risk of bias in the primary studies (internal validity). In this regard, none of the studies met all the methodological quality criteria. Five of them [19, 31, 32, 34, 35] were considered to have a high risk of bias, four were considered to have a low risk [17, 20, 28, 36], and one was rated as unclear [35]. The main issue was related to patient selection. Expert recommendations in diagnostic test accuracy (DTA) studies suggest that including patients selected as cases and controls should be avoided, as this can bias the diagnostic accuracy of a test. As shown in **Table 1**, only 4 studies included a consecutive sample of patients. Regarding applicability, or the extent to which the results can be generalized to other patients and settings, all studies were considered to have a low risk of bias. **Table 2** provides a summary of the different aspects analyzed in the risk of bias assessment.

TABLE 2
Evaluacion de riesgo de sesgo

STUDY	RISK OF BIAS				APPLICABILITY ISSUES		
	Patient selection	Index test	Reference standard	Flow and synchronization	Patient selection	Index test	Reference standard
Study 1	☺	☺	☺	☺	☺	☺	☺
Study 2	☺	☺	?	☺	☺	☺	☺
Study 3	☺	☺	☹	?	☺	☺	☺
Study 4	☺	☺	?	☺	☺	☺	?
Study 5	?	☺	?	?	☺	☺	?
Study 6	?	☹	?	☺	?	☺	☺
Study 7	☹	?	☺	?	☺	☺	☺
Study 8	☹	?	☺	☹	?	☺	☺
Study 9	☺	☺	☺	?	☺	☺	☺
Study 10	☹	?	?	?	?	☺	☺

Referencias: ☺ Low risk ☹ High risk ? Unclear risk

RESULTS OF INDIVIDUAL STUDIES

Despite the significant heterogeneity among the studies, both the 3Q-TMD test and the FAI had very good diagnostic accuracy. For TMD diagnosis, acceptable levels of sensitivity and specificity have been suggested to be at least 70% and 95%, respectively [14].

In contrast to diagnostic studies, screening studies focus on sensitivity, even at the expense of decreasing specificity, as the goal is to detect all suspicious individuals. The two studies on the 3Q-TMD test used the DC-TMD reference test and obtained the following diagnostic accuracy results: In Lovgren's study of the 3Q-TMD test in a primary setting, sensitivity was 0.81 (CI: 0.73–0.87), and specificity was 0.79 (CI: 0.73–0.83). In a secondary setting, sensitivity reached 0.96 (CI: 0.92–0.98), while specificity was 0.34 (CI: 0.28–0.40).

Regarding the studies that examined the FAI test, a high level of heterogeneity was observed. The results are categorized by the original version and the abbreviated version. Five studies on the original version [19, 31, 32, 35, 36] were identified, with most showing very good sensitivity and specificity data.

The studies on the short form of the FAI consist of three papers published between 2018 and 2021. The studies by Ujin-Yap and Pires [17,34] demonstrated very good sensitivity and specificity (0.95–0.93 for the Yap study, and 0.86–0.95 for the Pires study). The study by Zagalaz-Anula [37] showed acceptable sensitivity and specificity (0.78–0.79). Table 3 shows the results for sensitivity, specificity, and predictive values.

TABLE 3

Study	Year	Author	Sensitivity	CI 95%	Specificity	CI 95%	PPV	NPV
3Q-TMD Studies								
Study 1	2016	Lövgren	0.81	0.73-0.87	0.79	0.73-0.83	0.30	0.97
Study 2	2018	Lövgren	0.96	0.92-0.98	0.34	0.28-0.40	0.59	0.90
FAI Studies								
Study 3	2020	Stasiak	0.97	0.93-0.99	0.26	0.18-0.35	0.85	0.68
Study 4	2020	Kaynak	0.94	0.86- 0.96	0.83	0.75-0.93	0.92	0.85
Study 5	2019	Zhang	0.96	0.94- 0.97	0.72	0.59-0.83	0.97	0.62
Study 6	2014	Dos Santos	0.86	0.78- 0.91	0.92	0.83-0.96	0.94	0.84
Study 7	2021	Yap	0.95	0.93- 0.96	0.88	0.78-0.95	0.99	0.52
Abbreviated FAI Studies								
Study 8	2021	Zagalaz Anula	0.78	0.64-0.88	0.79	0.65-0.89	0.78	0.79
Study 9	2021	Yap	0.95	0.93-0.96	0.93	0.86-0.97	0.99	0.42
Study 10	2018	Pires	0.86	0.75-0.9	0.95	0.87-0.99	0.94	0.99

RESULTS SYNTHESIS

Despite the limited number of articles, a high degree of heterogeneity was observed among the analyzed studies due to variability in sample characteristics, methodological differences, and risk of bias. The meta-analysis results showed excellent sensitivity values for the FAI and 3Q-TMD tests: 0.92 (CI: 0.88–0.95). Specificity values demonstrated an accuracy close to 80%: 0.79 (CI: 0.63–0.90).

META-ANALYSIS

Sensitivity: In general, all studies showed very high sensitivity when using a fixed-effects model: 0.94 (CI: 0.93–0.95). Under a random-effects model, sensitivity was slightly lower, with a wider fluctuation range: 0.92 (CI: 0.88–0.95). No significant differences were observed between the index test (FAI) and the comparator test (3Q-TMD) (Figure 2).

Specificity: TMD screening tests showed lower specificity values than sensitivity, with greater heterogeneity among the results. Using a random-effects model, the meta-analysis indicated a specificity of 0.79 (CI: 0.63-0.90) (Figure 3).

Additional Measures:

Diagnostic Odds Ratio (DOR):

The overall DOR values for the set of tests were 3.81 (CI: 3.10–4.53). Based on the table of point estimates for the Diagnostic Odds Ratio, the analyzed tests demonstrated high sensitivity and specificity in terms of diagnostic accuracy, making them effective screening tools (Figure 4)

Proportional Hazards Model Approach: The proportional hazards model was used to estimate the performance of the tests on a receiver operating characteristic (ROC) curve. Under the homogeneity model, the overall results for the FAI and 3Q-TMD tests yielded an estimated area under the curve (AUC) of 0.95 (CI: 0.97–0.92). Under the heterogeneity model, the results were slightly lower but still demonstrated very good performance, with an AUC of 0.94 (CI: 0.97–0.91) (Figure 5).

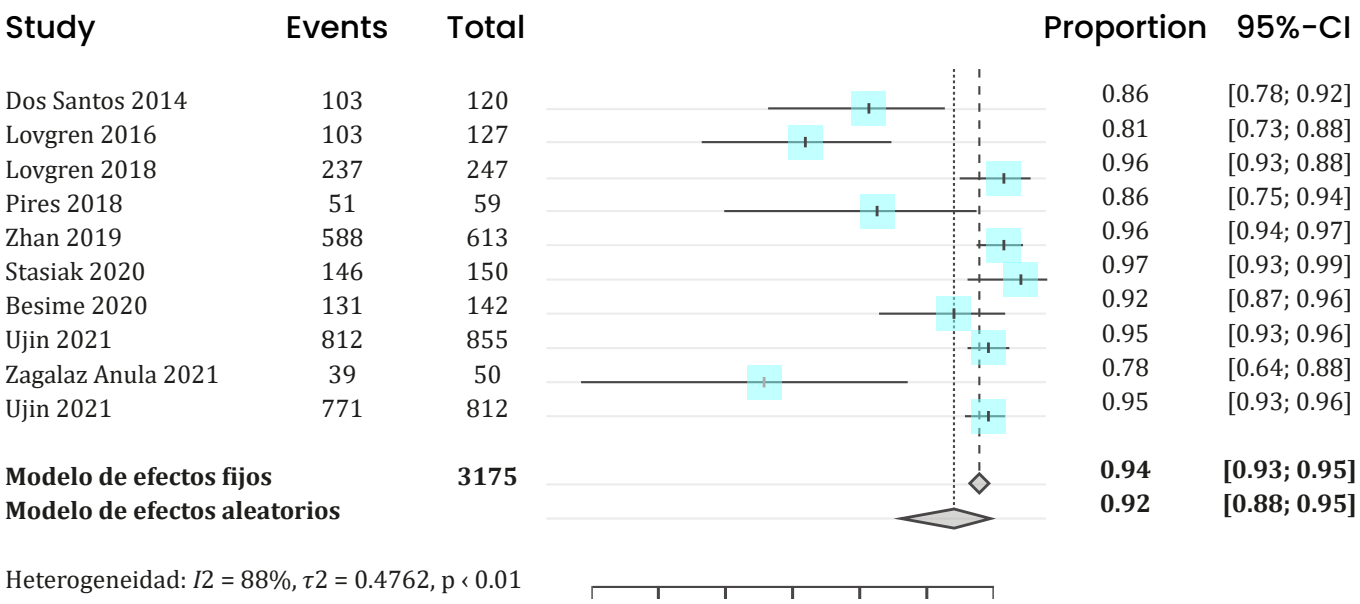

Figure 2 Forest plot of sensitivity

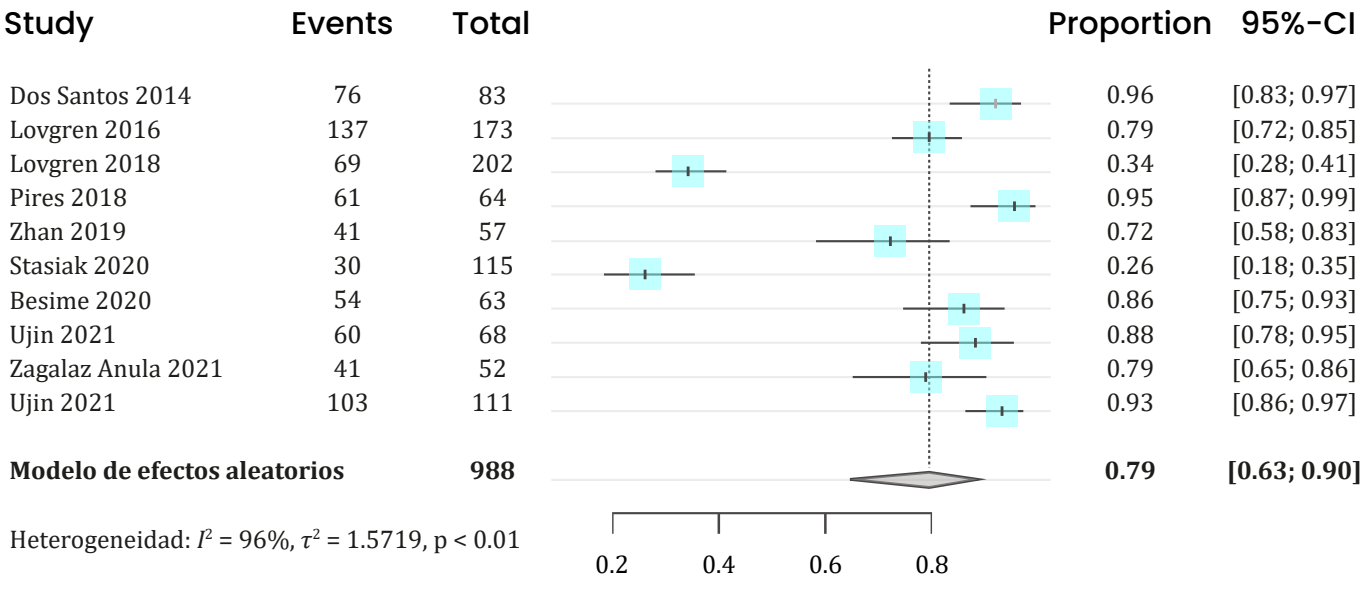

Figure 3 Forest plot of specificity.

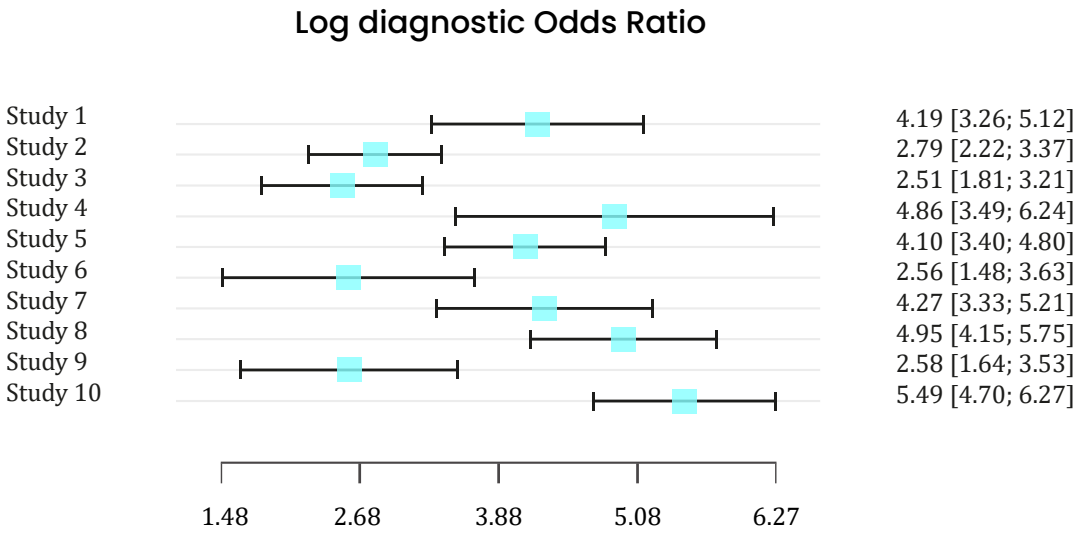

Figure 4 Forest plot of the Diagnostic Odds Ratio.

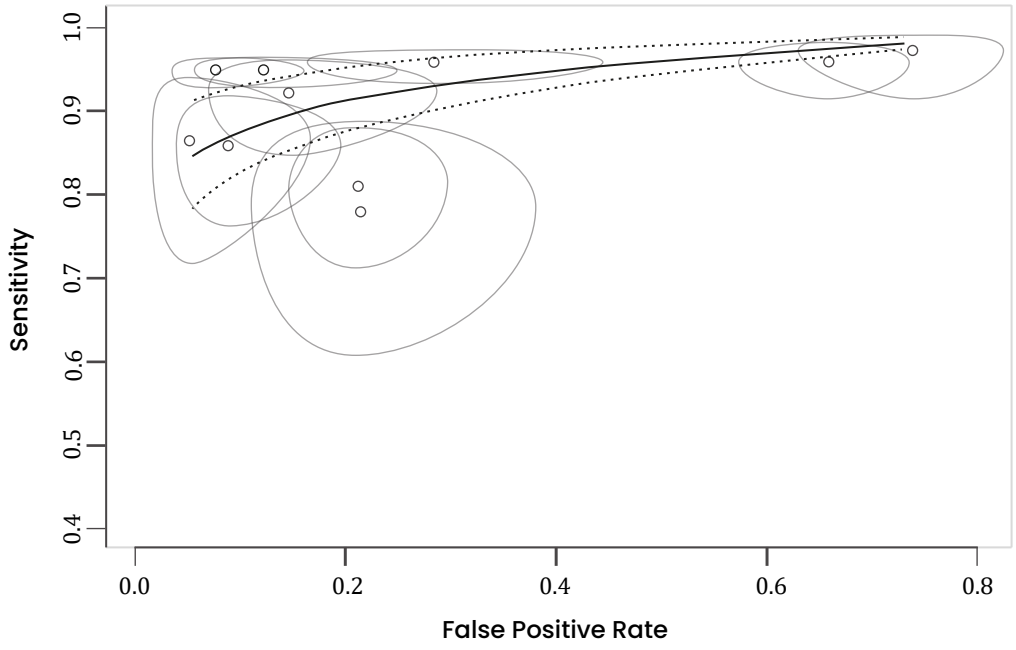

Figure 5 Area under the curve.

Discussion

Early diagnosis is critical in medical care as it defines the condition and confirms the patient's suffering. At times, as in the case of TMD, making a diagnosis can be challenging, but screening tools can help facilitate this process.

This study compared two simplified screening tests based on their accuracy against a reference diagnostic test (gold standard) to evaluate the feasibility of using them as screening tools.

REFERENCE TESTS

The RDC/TMD was developed in 1992 for research purposes only. Later, in 2014, the DC/TMD expanded its use

to clinical settings. These diagnostic tools are intended to establish reliable, standardized, and validated criteria for diagnosing TMD subtypes, as one of the main methodological issues in correlational research is the precise definition of the criteria applied [14]. Criterion validity for the screening tests was established in relation to the DC-TMD. Although the DC/TMD is reliable and valid, its routine use for clinical triage of TMD is not practical, as its evaluation protocol is time-consuming and requires proper interpretation of its complex algorithms [17]. In turn, screening tests for TMD detection provide a quick and simple way to determine which patients would benefit from a specific diagnosis.

It is worth noting that most of the screening studies in this review used the DC-TMD as the reference test. While the DC-TMD is the updated version of the RDC-TMD criteria, no significant difference in diagnostic accuracy was observed when the studies were analyzed through meta-analysis.

CRITERION VALIDITY OF THE 3Q-TMD TEST

The criterion validity of the 3Q-TMD test was established against the DC-TMD reference standard using two patient samples: one from the general population and the other from patients attending a specialized center. When the positive responses to the 3Q-TMD questions were compared with the reference test, a substantial difference between the settings was observed. However, the main reason for this difference may be due to the DC-TMD symptom questionnaire [16].

Upon examining the results in each setting, two key differences emerged. First, the time frame: although both questionnaires are based on reported symptoms, the 3Q-TMD test covers a one-week period, while the reference test relies on symptoms perceived over the past 30 days. Second, the phrasing of the questions: the first question in the 3Q-TMD questionnaire focuses on reported pain, and the second on functional pain. However, to qualify for a DC pain/TMD diagnosis, the criteria require the pain to be caused or modified by function. The observed differences in the sensitivity and specificity of the 3Q-TMD test may be attributed to these two factors. Nevertheless, the predictive values were high, particularly the negative predictive values, indicating that these questions are excellent for ruling out a TMD diagnosis.

The question related to functional disorders deserves a separate paragraph. There is ongoing discussion about the prognosis of joint sounds and the clinical evaluation possibilities for intra-articular TMD [38]. The DC/TMD has shown moderate to poor validity for intra-articular TMD. In this sense, the reference standard would perform better as a screening tool than as a diagnostic instrument [39].

Recently, the validity of the RDC-TMD and DC-TMD clinical protocols for evaluating intra-articular disorders has been questioned [40]. Imaging studies such as MRI, CT, and even arthrography have been proposed as more accurate alternatives. These methods, however, would only be justified if their results could alter the therapeutic protocol. It is important to note that even in asymptomatic individuals, disc displacements may be observed on MRI in approximately 30% of the population [41,42]. Unlike the two pain questions, the results of the intra-articular disorders question showed varying utility depending on the setting in which it was applied. When applied to the general population, it was useful for ruling out the absence of intra-articular TMD. However, when applied in specialized settings, its high positive predictive value indicated a probable intra-articular TMD. In summary, while the validity of the intra-articular TMD question ranges from fair to moderate, its high specificity makes it very useful for screening, particularly for ruling out dysfunction when the results are negative.

CRITERION VALIDITY OF THE FAI TEST

Unlike the 3Q-TMD test, the Fonseca Anamnestic Index identifies individuals with TMD from a score. The FAI was originally designed to determine the severity of symptoms. All FAI studies have demonstrated high accuracy in detecting pain-related and intra-articular TMDs, with this study reporting an area under the curve ranging from 0.93 to 0.98 across various observations [21,32,36]. Studies analyzing the FAI established scores derived from questions that included individuals with pain, those with intra-articular disorders, and the sum of all symptoms. In this regard, the classification was similar to that used in the 3Q-TMD test. The FAI studies, however, differed in their cutoff points, as each study defined different values to achieve the highest accuracy relative to the reference test. These variations may be explained by the heterogeneity of populations and methods used. FAI cutoff points for ruling out individuals without TMD ranged from 0 to 20, with higher scores correlating with increased symptom intensity. While uncertainties remain regarding the utility of assigning a degree of symptom severity, various studies have confirmed the FAI's ability to identify TMD. Overall, the FAI appears to be highly sensitive for detection, though its specificity is relatively low. The lower specificity observed with the FAI may be attributed to the inclusion of items unrelated to TMD, such as headaches, neck pain, parafunctional habits, malocclusion, and emo-

tional stress. This prompted further investigation into the questionnaire's dimensionality and psychometric properties [43]. The study confirmed the FAI's multidimensionality, identifying a primary five-item dimension, which resulted in the development of the five-question short-form FAI. Research on this abbreviated version showed improved accuracy, with an area under the curve of 0.97 and significantly higher specificity (95.5%) relative to the reference test.

ANALYSIS OF THE FAI AND 3Q-TMD TESTS

The results of this study demonstrated that, when comparing both tests, the larger number of published studies on the FAI test allows its diagnostic accuracy results to be considered reliable, despite the significant heterogeneity among studies. In contrast, only two studies were found on the 3Q-TMD test. In this case, the reliability of its results is supported by the low risk of bias assessed in both studies. It is important to note that, despite the heterogeneity in methods and settings, the FAI and 3Q-TMD screening tests have shown very good diagnostic accuracy. Only two studies reported low specificity: Lovgren's study on the 3Q-TMD test in a specialized setting (specificity 0.34) and Stasiak's study on the FAI test (specificity 0.26), also conducted in a secondary setting. Regarding Lovgren's study, performed in a specialized orofacial pain clinic, the high false positive rate could be attributed to the greater frequency of facial pain unrelated to TMD in such settings compared to primary care environments. It is important to emphasize that the DC/TMD was selected due to its reliability and validity for the most common TMD diagnoses. However, numerous additional pain conditions can affect normal jaw function, such as neuropathic pain, atypical odontalgia, fibromyalgia, and cervical pain. The prevalence of both painful and non-painful conditions is expected to be much higher in specialized orofacial pain clinics than in general population-based clinics. Similarly, it is logical that rarer conditions would also have a much higher prevalence in specialized settings [44]. Therefore, affirmative responses to the 3Q-TMD test in a specialized clinic may be associated with a TMD diagnosis but could also reflect several differential diagnoses. Ultimately, this may explain the increase in false positives and the consequent decrease in specificity.

The same hypothesis could be applied to the results of the Stasiak study. However, there are valid reasons to attribute this difference to the risk of bias. In particular, the Stasiak study does not clearly specify how the reference test was applied nor how patient flow and timing were managed. The reference standard used in these studies relies on a diagnostic system based on strict criteria, with both the clinical history and examination being meticulously structured. The data are subsequently processed using predefined algorithms, resulting in a probable diagnosis. Additionally, the examiner requires appropriate training. Therefore, improper handling of the reference test could explain the discrepancy in results. Supporting this explanation is the study by Yap, which also used the FAI and was conducted in a secondary setting. However, it reported a much higher specificity (0.88). Based on the available information, it is not possible to determine the reasons behind the differences in specificity observed for the FAI test in secondary settings. Overall, these findings highlight the need for further studies of this nature to elucidate such differences. It is important to clarify, however, that when these tests are regarded as screening tools rather than diagnostic instruments, a loss in specificity does not constitute a limitation. In fact, it is expected that screening tests do not exclude any potential patients with the condition, even at the "expense" of increased false positives. In this sense, unlike diagnostic studies, screening studies prioritize sensitivity over specificity, as it is crucial to detect all individuals who may be at risk [45].

In terms of sensitivity, the results of almost all studies demonstrated a very good ability to detect TMD, with the exception of the Zagalaz study, which showed a slight decrease (Sensitivity 0.78). However, this result may be questioned, as it originates from the study with the highest risk of bias. Despite this, both the FAI and the 3Q-TMD showed good performance in identifying individuals with TMD.

While there may be uncertainties regarding specific aspects of the results, the lack of well-established tools for TMD detection underscores that this evidence, although not definitive, represents the best currently available and supports promoting the use of these instruments. Since the 3Q-TMD test has been validated in its original language and the FAI in only three other languages, researchers should be encouraged to validate either of these tools in their respective languages.

This systematic review has some limitations. First, only a small number of studies were identified, with variable results regarding the diagnostic accuracy of each test and significant methodological heterogeneity. Second, a high risk of bias was observed.

FUTURE GUIDANCE

As a guideline for future studies, it is recommended that researchers take special care in patient selection. Specifically, the inclusion of patients as "cases and controls" should be avoided, as this may bias the results, according to specialists in diagnostic accuracy testing. Notably, among

all the studies analyzed, this was the aspect that had the greatest influence on the risk of bias assessment. The heterogeneity and small number of studies found indicate that TMD screening tests are only beginning to be recognized as necessary tools in the field of craniofacial pain of non-odontogenic origin. This observation is significant in this review because, despite some uncertainties in the results, the lack of validated tools for TMD detection highlights that this evidence, while not conclusive, represents the best currently available and supports promoting their use. Given that these instruments have been validated in their original language and, in the case of the FAI, in only three other languages, researchers are encouraged to validate them in their own language.

Conclusions

The FAI and 3Q-TMD tests are questionnaires with short items and are simple, practical tools that can be routinely applied in clinical settings without disrupting daily activities.

Based on the results of this systematic review and meta-analysis, it can be concluded that both the FAI and the 3Q-TMD are highly sensitive instruments for detecting TMD in patients. Both screening tests make it easier to identify individuals with TMD. In this context, patients detected early could benefit from timely diagnosis and treatment, avoiding prolonged searches for a diagnosis and reducing treatment costs by preventing the condition from becoming chronic.

To the best of our knowledge, this is the first systematic review with a meta-analysis of TMD screening tests. Future validation studies will provide new data and more reliable information on the ability of these tests to detect TMD.

REFERENCES

1. International Classification of Orofacial Pain, 1st edition (ICOP). Cephalalgia. 2020 Feb 1;40(2):129–221.

2. Valesan LF, Da-Cas CD, Réus JC, Denardin ACS, Garanhani RR, Bonotto D, et al. Prevalence of temporomandibular joint disorders: a systematic review and meta-analysis. Vol. 25, Clinical Oral Investigations. Springer Science and Business Media Deutschland GmbH; 2021. p. 441–53.

3. List T, Jensen RH. Temporomandibular disorders: Old ideas and new concepts. Vol. 37, Cephalalgia. SAGE Publications Ltd; 2017. p. 692–704.

4. Dahlstrm L, Carlsson GE. Temporomandibular disorders and oral health-related quality of life. A systematic review. Acta Odontol Scand. 2010;68(2):80–5.

5. Wänman A, Wigren L. Need and demand for dental treatment a comparison between an evaluation based on an epidemiologic study of 35-, 50-, and 65-year-Olds and performed dental treatment of matched age groups. Acta Odontol Scand. 1995;53(5):318–24.

6. Ameer Al-Jundi M, John MT, Setz JM, Szentpétery A, Kuss O. Meta-analysis of treatment need for temporomandibular disorders in adult nonpatients. J Orofac Pain. 2008 Mar;22(2):97–107.

7. Magnusson T, Egermark I, Carlsson GE. A Longitudinal Epidemiologic Study of Signs and Symptoms of Temporomandibular Disorders from 15 to 35 Years of Age. J Orofac Pain. 2000;14(4):310–9.

8. Lundh H, Westesson PL, Kopp S. A three-year follow-up of patients with reciprocal temporomandibular joint clicking. Oral Surgery, Oral Medicine, Oral Pathology. 1987;63(5):530–3.

9. Alex White B, Williams LA, Leben JR. Health Care Utilization and Cost among Health Maintenance Organization Members with Temporomandibular Disorders. J Orofac Pain. 2001;15(2):158–69.

10. National Institute of Dental and Craniofacial Research. Facial Pain. http://www.nidcr.nih.gov/DataStatistics/FindDataByTopic/ FacialPain/ [Internet]. [cited 2022 May 17]. Available from: https://www.nidcr.nih.gov/research/data-statistics/facial-pain

11. Stowell AW, Wildenstein L, Gatchel RJ. Cost-effectiveness of treatments for temporomandibular disorders: Biopsychosocial intervention versus treatment as usual. Journal of the American Dental Association [Internet]. 2007;138(2):202–8. Available from: http://dx.doi.org/10.14219/jada.archive.2007.0137

12. Durham J, Steele J, Moufti MA, Wassell R, Robinson P, Exley C. Temporomandibular disorder patients' journey through care. Community Dent Oral Epidemiol. 2011;39(6):532–41.

13. Macfarlane GJ. The epidemiology of chronic pain. Pain [Internet]. 2016 Jun 30 [cited 2022 May 31];157(10):2158–9. Available from: https://pubmed.ncbi.nlm.nih.gov/27643833/

14. Dworkin SF, LeResche L. Research diagnostic criteria for temporomandibular disorders: review, criteria, examinations and specifications, critique. J Craniomandib Disord. 1992;6(4):301–55.

15. Schiffman E, Ohrbach R, Truelove E, Look J, Anderson G, Goulet JP, 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 Group†. J Oral Facial Pain Headache. 2014;28(1):6–27.

16. Lövgren A. Recognition of Temporomandibular Disorders. UMEA; 2017.

17. Yap AU, Zhang MJ, Lei J, Fu KY. Diagnostic accuracy of the short-form Fonseca Anamnestic Index in relation to the Diagnostic Criteria for Temporomandibular Disorders. Journal of Prosthetic Dentistry. 2021;

18. Magnusson T, Egermark I, Carlsson GE. Treatment received, treatment demand, and treatment need for temporomandibular disorders in 35-year-old subjects. Cranio [Internet]. 2002 [cited 2022 May 31];20(1):11–7. Available from: https://pubmed.ncbi.nlm.nih.gov/11831338/

19. Stasiak G, Maracci LM, de Oliveira Chami V, Pereira DD, Tomazoni F, Bernardon Silva T, et al. TMD diagnosis: Sensitivity and specificity of the Fonseca Anamnestic Index. Cranio [Internet]. 2020 Oct 27 [cited 2021 Jun 1];1–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/33108257

20. Lovgren A, Visscher CM, Haggman-Henrikson B, Lobbezoo F, Marklund S, Wanman A. Validity of three screening questions (3Q/TMD) in relation to the DC/TMD. J Oral Rehabil. 2016;43(10):729–36.

21. Berni KC dos S, Dibai-Filho AV, Rodrigues-Bigaton D. Accuracy of the Fonseca anamnestic index in the identification of myogenous temporomandibular disorder in female community cases. J Bodyw Mov Ther. 2015 Jul 1;19(3):404–9.

22. Pastore GP, Goulart DR, Pastore PR, Prati AJ, de Moraes M. Comparison of instruments used to select and classify patients with temporomandibular disorder. Acta Odontológica Latinoamericana [Internet]. 2018 [cited 2023 Aug 10];31(1):16–22. Available from: http://www.scielo.org.ar/scielo.php?script=sci_arttext&pid=S1852-48342018000100003&lng=es&nrm=iso&tlng=en

23. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (prisma-p) 2015: Elaboration and explanation. Vol. 349, BMJ (Online). BMJ Publishing Group; 2015.

24. Salameh JP, Bossuyt PM, McGrath TA, Thombs BD, Hyde CJ, MacAskill P, et al. Preferred reporting items for systematic review and meta-analysis of diagnostic test accuracy studies (PRISMA-DTA): Explanation, elaboration, and checklist. Vol. 370, The BMJ. BMJ Publishing Group; 2020.

25. INFORM International Network for Orofacial Pain and Related Disorders Methodology [Internet]. 2023 [cited 2023 Mar 19]. Available from: .buffalo.edu/rdc-tmdinternational/

26. Health Information Research Unit - HIRU ~ Search Strategies for MEDLINE in Ovid Syntax and the PubMed translation [Internet]. 2022 [cited 2022 May 17]. Available from: https://hiru.mcmaster.ca/hiru/HIRU_Hedges_MEDLINE_Strategies.aspx

27. Ciapponi A. QUADAS-2: instrumento para la evaluación de la calidad de estudios de precisión diagnóstica QUADAS-2: an instrument for the evaluation of the quality of diagnostic precision studies. 2015;22–30. Available from: http://www.bris.ac.uk/quadas

28. Lovgren A, Parvaneh H, Lobbezoo F, Haggman-Henrikson B, Wanman A, Visscher CM. Diagnostic accuracy of three screening questions (3Q/TMD) in relation to the DC/TMD in a specialized orofacial pain clinic. Acta Odontol Scand. 2018;76(6):380–6.

29. Kaynak BA, Taş S, Salkın Y. The accuracy and reliability of the Turkish version of the Fonseca anamnestic index in temporomandibular disorders. Cranio [Internet]. 2020 Aug 25 [cited 2021 Jun 1];1–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/32840464

30. Zhang M juan, Yap AUJ, Lei J, Fu KY. Psychometric evaluation of the Chinese version of the Fonseca anamnestic index for temporomandibular disorders. J Oral Rehabil. 2020 Mar 1;47(3):313–8.

31. Berni KC dos S, Dibai-Filho AV, Rodrigues-Bigaton D. Accuracy of the Fonseca anamnestic index in the identification of myogenous temporomandibular disorder in female community cases. J Bodyw Mov Ther. 2015 Jul 1;19(3):404–9.

32. Yap AU, Zhang MJ, Lei J, Fu KY. Accuracy of the Fonseca Anamnestic Index for identifying pain-related and/or intra-articular Temporomandibular Disorders. Cranio - Journal of Craniomandibular Practice. 2021;

33. Zagalaz-Anula N, Sanchez-Torrelo C, Acebal-Blano F. The Short Form of the Fonseca Anamnestic Index for the Screening of Temporomandibular Disorders_ Validity and Reliability in a Spanish-Speaking Population _ Enhanced Reader. J Clin Med. 2021;

34. Pires PF, de Castro EM, Pelai EB, de Arruda ABC, Rodrigues-Bigaton D. Analysis of the accuracy and reliability of the Short-Form Fonseca Anamnestic Index in the diagnosis of myogenous temporomandibular disorder in women. Braz J Phys Ther [Internet]. 2018;22(4):276–82. Available from: https://www.embase.com/search/results?subaction=viewrecord&id=L624196653&from=export

35. Zhang M juan, Yap AUJ, Lei J, Fu KY. Psychometric evaluation of the Chinese version of the Fonseca anamnestic index for temporomandibular disorders. J Oral Rehabil. 2020 Mar 1;47(3):313–8.

36. Kaynak BA, Taş S, Salkın Y. The accuracy and reliability of the Turkish version of the Fonseca anamnestic index in temporomandibular disorders. Cranio - Journal of Craniomandibular Practice. 2020;

37. Zagalaz-Anula N, María Sánchez-Torrelo C, Acebal-Blanco F, Alonso-Royo R, Javier Ibáñez-Vera A, Obrero-Gaitán E, et al. Clinical Medicine The Short Form of the Fonseca Anamnestic Index for the Screening of Temporomandibular Disorders: Validity and Reliability in a Spanish-Speaking Population. J Clin Med. 2021;10:5858.

38. Giraudeau A, Jeany M, Ehrmann E, Déjou J, Ouni I, Orthlieb JD. Disc displacement without reduction: a retrospective study of a clinical diagnostic sign. Cranio - Journal of Craniomandibular Practice. 2017 Mar 4;35(2):86–93.

39. Schiffman E, Ohrbach R. Executive summary of the Diagnostic Criteria for Temporomandibular Disorders for clinical and research applications. Journal of the American Dental Association. 2016 Jun 1;147(6):438–45.

40. Pupo YM, Pantoja LLQ, Veiga FF, Stechman-Neto J, Zwir LF, Farago PV, et al. Diagnostic validity of clinical protocols to assess temporomandibular disk displacement disorders: a meta-analysis. Vol. 122, Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology. Mosby Inc.; 2016. p. 572–86.

41. Katzberg RW, Westesson PL, Tallents RH, Drake CM, Davis R. Orthodontics and temporomandibular joint internal derangement. Vol. 109, Am J Orthod Dentofac Orthop. 1996.

42. Salé H, Bryndahl F, Isberg A. Temporomandibular joints in asymptomatic and symptomatic nonpatient volunteers: A prospective 15-year follow-up clinical and MR imaging study. Radiology. 2013 Apr;267(1):183–94.

43. Rodrigues-Bigaton D, de Castro EM, Pires PF. Factor and Rasch analysis of the Fonseca anamnestic index for the diagnosis of myogenous temporomandibular disorder. Braz J Phys Ther. 2017 Mar 1;21(2):120–6.

44. Manfredini D, Guarda-Nardini L, Winocur E, Piccotti F, Ahlberg J, Lobbezoo F. Research diagnostic criteria for temporomandibular disorders: A systematic review of axis i epidemiologic findings. Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology and Endodontology. 2011;112(4):453–62.

45. Talavera JO, Wacher-Rodarte NH, Rivas-Ruiz R, Nacional CM, Xxi S. Estudios de proceso (prueba diagnóstica). Vol. 49, Rev Med Inst Mex Seguro Soc. 2011.

Data availability

The entire dataset supporting the results of this study has been published in the article itself.

Conflict of Interest Declaration

The authors declare that there are no conflicts of interest.

Funding source

The author declares not having funding source.

Authorship Contribution and Collaboration Statement

NAME AND LAST NAME	ACADEMIC COLLABORATION													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Hugo Daniel Rivera	x			x	x	x	x	x		x				x
Claudia Ivonne Rodriguez				x	x	x	x		x					
Ramón Álvarez-Vaz			x		x	x		x			x		x	x
Marcelo Kreiner	x	x		x		x		x				x	x	

1. Project Administration
2. Funding Acquisition
3. Formal Analysis
4. Conceptualization
5. Data Curation
6. Writing - Review & Editing
7. Research
8. Methodology
9. Resources
10. Writing-Original Draft Preparation
11. Software
12. Supervision
13. Validation
14. Visualization

Acceptance Note:
This article was approved by the journal editor MsC. Dra. Natalia Tancredi