Facultad de Odontología
95 AÑOS

Odontoestomatología
DOI: 10.22592/ode2025n45e338

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

ACTUALIZACIÓN

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

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


## Resumen

El objetivo de esta revisión fue evaluar la validez de dos instrumentos de detección de trastornos témporomandibulares, 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.



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

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


Recibido: 10/Abr/2024
Aceptación: 05/Nov/2024

CC BY

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


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.

## Abstract

The aim of this review was to assess the validity of two screening instruments for temporomandibular disorders (the IAF 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 IAF and 3Q-TMD instruments are valid for detection of individuals with TMD.


**Keywords:** Craniomandibular Disorders. Craniomandibular Dysfunctions. Temporomandibular joint disorders, triage, orofacial pain.


## 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.

# Introducción

Los trastornos témporomandibulares (TTM) son afecciones que se caracterizan clínicamente por dolor y/o disfunción en los músculos masticatorios, cervicales, de la cabeza, las articulaciones témporomandibulares (ATM) y estructuras adyacentes [1].

Los TTM son un problema de salud pública significativo que afecta aproximadamente al 30% de la población general [2] y se considera la causa más común de dolor crónico de origen no dental en el área orofacial [3]. El dolor relacionado con los TTM puede afectar la capacidad del individuo para realizar sus actividades diarias, el funcionamiento psicosocial y su calidad de vida [4]. Sin embargo, aun cuando los TTM tienen un impacto negativo en la vida diaria, estas condiciones parecen ser en gran medida no detectadas o descuidadas en la atención dental de rutina. Esto se ve reflejado por la discrepancia entre la necesidad estimada de tratamiento y el tratamiento efectivamente realizado [5,6]. Solo cuando el dolor es muy intenso o se hace crónico los pacientes tienden a buscar tratamiento [7,8]. Los estudios realizados sobre las implicancias económicas y laborales de las entidades dolorosas crónicas como las cefaleas o el dolor facial por TTM son muy elevadas. Se estima que los pacientes con TTM generan un 50% más de costos medios de consumos de fármacos y consultas profesionales [9]. Se ha calculado que solo en EE. UU. el gasto del tratamiento de estas afecciones en la última década se ha duplicado, estimándose en unos 4.000 millones de dólares anuales [10]. Diferentes investigaciones indican que la intervención temprana de estos trastornos tiene una alta tasa de éxito y conllevan menores costos de tratamiento a largo plazo [11–13].

La utilización rutinaria de pruebas de detección de TTM en la práctica odontológica permitiría que estos pacientes accedan a un diagnóstico y tratamiento oportuno, lo que mejoraría su calidad de vida, disminuyendo además los costos de tratamiento.

## PRUEBAS DE DIAGNÓSTICO EN TTM

Existen varios sistemas de diagnóstico para el dolor orofacial causado por TTM [14,15]. Los Criterios de Diagnóstico para TTM, (DC/TMD) y su versión previa, el RDC / TMD, son métodos de diagnóstico estrictamente definidos para las afecciones más comunes de TTM como

mialgia, artralgia, dolor miofascial, trastornos articulares degenerativos y trastornos funcionales de la articulación témporomandibular. Su sistema de evaluación, que separa los aspectos psicosociales de los aspectos físicos, ha sido considerado un cambio de paradigma dentro del campo del dolor orofacial [16].

Los protocolos diagnósticos RDC y DC/TMD han sido utilizados, en los últimos años, como prueba de referencia o "gold standard" para evaluar las pruebas de screening en la detección de TTM. Si bien el DC/TMD es fiable y válido, su uso rutinario para el triaje clínico de TTM no es práctico dado que su protocolo de evaluación consume mucho tiempo y requiere la adecuada interpretación de sus complejos algoritmos [17]. En este sentido, las herramientas de detección de TTM deben ser económicas, sencillas, eficientes y precisas.

## PRUEBAS DE SCREENING PARA LA DETECCIÓN DE TTM

Se ha sugerido que la detección temprana permitiría prevenir los cuadros crónicos [13]. A su vez, la mayoría de los pacientes con diagnóstico de TTM se beneficiarían con tratamientos conservadores y únicos [18].

Se han validado varios métodos diagnósticos para la detección de TTM. Los más estudiados han sido el screening de dolor por TTM de la prueba RDC-TMD, el Índice Anamnésico de Fonseca (IAF) [19] y el Test 3Q-TMD [20]. De estas tres pruebas, el screening RDC-TTM ha demostrado validez para detectar individuos con dolor potencial por TTM, pero no incluye aspectos funcionales no dolorosos [16]. El IAF es un cuestionario de 10 preguntas en su forma original y de 5 preguntas en su forma corta (IAFC) que permite detectar la existencia de TTM. Las preguntas poseen un puntaje que indican los grados de severidad del trastorno. El IAF presentó resultados consistentes con otros instrumentos para detectar TTM, incluido el cuestionario de la academia americana de dolor orofacial y se evaluó contra el RDC / TMD y DC / TMD [21,22].

Por su parte, el test 3Q-TMD es más breve y se compone de un cuestionario de 3 preguntas, dos de las cuales se orientan sobre el dolor facial y la tercera se enfoca en la existencia de bloqueo mandibular durante la función. Dado que los pacientes con TTM parecen no ser detectados y tratados oportunamente por la odontología, hemos centrado este trabajo en los 2 métodos de detección que permitirían identificar los TTM más comunes, tanto dolorosos como funcionales. Por lo tanto, esta revisión sistemática se ha enfocado en comparar estos dos instrumentos a través de la siguiente pregunta: ¿Cuál es la validez (especificidad y sensibilidad: precisión) de los instrumentos disponibles para uso clínico y epidemiológico IAF y 3Q-TMD para la detección de TTM?

# Métodos

## PROTOCOLO Y REGISTRO

El protocolo de esta revisión fue documentado de acuerdo con la guía de la declaración PRISMA-P [23] y fue enviado para su inscripción en el Registro Internacional Prospectivo de Revisiones Sistemáticas (PROSPERO, Centre for Reviews and Dissemination, University of York; y National Institute for Health Research). Para elaborar esta revisión sistemática de pruebas diagnósticas se siguieron las recomendaciones de la lista de verificación PRISMA-DTA [24].

## CRITERIOS DE ELEGIBILIDAD

En esta revisión se incluyeron estudios de precisión diagnóstica para detección de TTM basados en el IAF como prueba índice y el test 3Q-TMD como comparador. Para ambas pruebas se examinaron los trabajos que tuvieran como gold standard las pruebas RDC y DC/TMD. No se aplicaron restricciones con respecto a la edad y el sexo de los participantes ni con respecto al idioma de publicación de los trabajos. Se admitieron tanto estudios con un diseño tipo "cohorte" como aquellos con un diseño "caso-control".

En cuanto a los sitios de recolección de muestra, se aceptaron estudios tanto de entornos primarios (centros odontológicos generales) como secundarios (centros especializados en dolor orofacial). Para establecer la condición objetivo, los pacientes de estos estudios respondieron a una evaluación clínica individual de TTM (partcularmente ruidos articulares, limitación del movimieto mandibular, dolor muscular, articular y/o dolor pre auricular) de acuerdo a los criterios establecidos por el consorcio INFORM [25].

Los criterios de exclusión abarcaron lo siguiente:

- Estudios que no utilizaron RDC/TMD (estudios publicados antes de 1992) o DC/TMD o estudios que modificaron la herramienta.
- Artículos con datos duplicados de otro estudio incluido.
- Estudios que no estuviesen enfocados exclusivamente en el diagnóstico de pacientes con TTM.
- Revisiones, cartas, libros, opinión de expertos e informes de casos.

## IDENTIFICACIÓN DE LAS FUENTES DE ESTUDIO

Se utilizó un filtro de búsqueda HIRU para estudios de precisión diagnóstica [26]. A continuación, se desarrolló una estrategia de búsqueda electrónica para PubMed y se adaptó para cada una de las siguientes bases de datos

bibliográficas: Web of Science, Science direct, Scopus y Scielo. También se buscó en bases de datos de literatura gris como Google Scholar y ProQuest One Academic.

El período de búsqueda abarcó desde el año 1992 hasta abril de 2022. Además, se realizaron búsquedas manuales en la lista de referencias de los estudios incluidos para identificar estudios relevantes adicionales. Se utilizó un gestor de referencias* para recopilar referencias y eliminar duplicados (*Mendeley®, Elsevier Ámsterdan, Paises Bajos).

## ESTRATEGIA DE BÚSQUEDA

Se utilizaron términos indexados y términos libres para localizar la investigación realizada sobre estudios de diagnóstico de TTM, así como su precisión diagnóstica. Se describe a continuación la estrategia de búsqueda utilizando el filtro para estudios de precisión diagnóstica [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]).

## SELECCIÓN DE ESTUDIOS

En la fase 1, dos autores (HDR y CIR) evaluaron de forma independiente los títulos y resúmenes de los estudios identificados aplicando los criterios de elegibilidad establecidos previamente. Cuando los artículos se consideraron elegibles para su inclusión, los revisores realizaron una lectura de texto completo en la fase 2. Finalmente se eliminaron los artículos que no respondían a los criterios de elegibilidad. Si bien se contaba con un tercer revisor (MK) para intervenir ante posibles desacuerdos, los mismos se resolvieron entre los autores en varias reuniones de consenso.

## PROCESO DE RECOPILACIÓN DE DATOS

El proceso de recolección de datos fue realizado por el primer revisor (HDR) y verificado por el segundo revisor (CIR) para asegurar la integridad de los contenidos. Los desacuerdos se consensuaron con el tercer revisor (MK). Se utilizó una planilla electrónica y se extrajeron los siguientes datos para cada estudio incluido:

Datos bibliométricos: año de publicación, país de origen, autor, revista donde fue publicada.

Datos metodológicos: número de participantes de cada estudio, edad, sexo, forma de cegamiento y randomización de los pacientes para las pruebas. Entorno sanitario donde se realizó el estudio (atención primaria o secundaria).

Datos estadísticos: se extrajeron datos como: tamaño muestral de cada grupo, datos de sensibilidad y especificidad, valores predictivos, OR diagnóstico y área bajo la curva ROC.

## RIESGO DE SESGO

El riesgo de sesgo fue evaluado de forma indepen diente por dos revisores (HDR y CIR). Se utilizó la herramienta QUADAS 2 para examinar el riesgo de sesgo y la aplicabilidad de los estudios [27]. La evaluación comprende cuatro dominios separados: selección de pacientes, prueba índice, estándar de referencia y flujo y tiempo de pacientes a través del estudio. Para comenzar con la aplicación de la herramienta se realizó previamente una prueba piloto para garantizar un consenso sobre la valoración del riesgo de sesgo entre ambos revisores. No se calculó una puntuación general resumida, pero para cada dominio, cualquier preocupación con respecto al sesgo y la aplicabilidad se calificó como “baja”, “alta” o “poco clara”. Las diferencias se resolvieron por consenso. Estos resultados se presentan en forma completa en un documento anexo.

## MEDIDAS DE PRECISIÓN DIAGNÓSTICA

Se utilizaron datos de las tablas de dos por dos para calcular la sensibilidad, especificidad y valores predictivos de cada estudio. Los resultados de los estudios individuales se presentaron gráficamente, trazando las estimaciones de sensibilidad y especificidad (y sus intervalos de confianza del 95 %). Se realizaron estimaciones adiciona-

les como el Odds Ratio diagnóstico y se utilizó un modelo de riesgos proporcionales para estimar el comportamiento de las pruebas sobre una curva de rendimiento diagnóstico (ROC), de este modo se pudo apreciar en un solo valor la capacidad de detección de las pruebas de detección de manera global.

### SÍNTESIS DE RESULTADOS

Los resultados se resumieron trazando estimaciones de sensibilidad y especificidad, para las pruebas índice y comparador (IAF y 3Q-TMD) en diagramas de bosque acoplados y en una matriz de características operativas del receptor (Curvas ROC).

### METAANÁLISIS

Para ambas pruebas a evaluar se estimaron los puntos operativos resumidos (sensibilidades y especificidades resumidas) bajo un intervalo de confianza del 95 %. A continuación, se estableció el odds ratio diagnóstico para obtener en un solo número la precisión diagnóstica para cada prueba. Finalmente se calculó en una curva de rendimiento diagnóstico (ROC) el área bajo la curva para ambas pruebas utilizando un modelo de riesgo proporcionales.

## Resultados

### SELECCIÓN DE ESTUDIOS

Se encontraron un total de 798 artículos. Se excluyeron 727 artículos debido a títulos no elegibles para los objetivos de la búsqueda. En esta etapa los motivos de exclusión de títulos obedecían principalmente a que los trabajos no correspondían a pruebas de detección. Después de seleccionar 71 estudios para la evaluación de títulos y resúmenes se eliminaron 49 trabajos repetidos (fase 1). Finalmente 22 estudios se consideraron elegibles para ser evaluados íntegramente. Después de la lectura del texto completo (fase 2), se excluyeron 12 estudios que no tenían como eje central la evaluación de pruebas de screening. Finalmente, se incluyeron 10 artículos para realizar la revisión sistemática y el metanálisis. Una visión general del proceso de selección se muestra en la **Figura 1**.

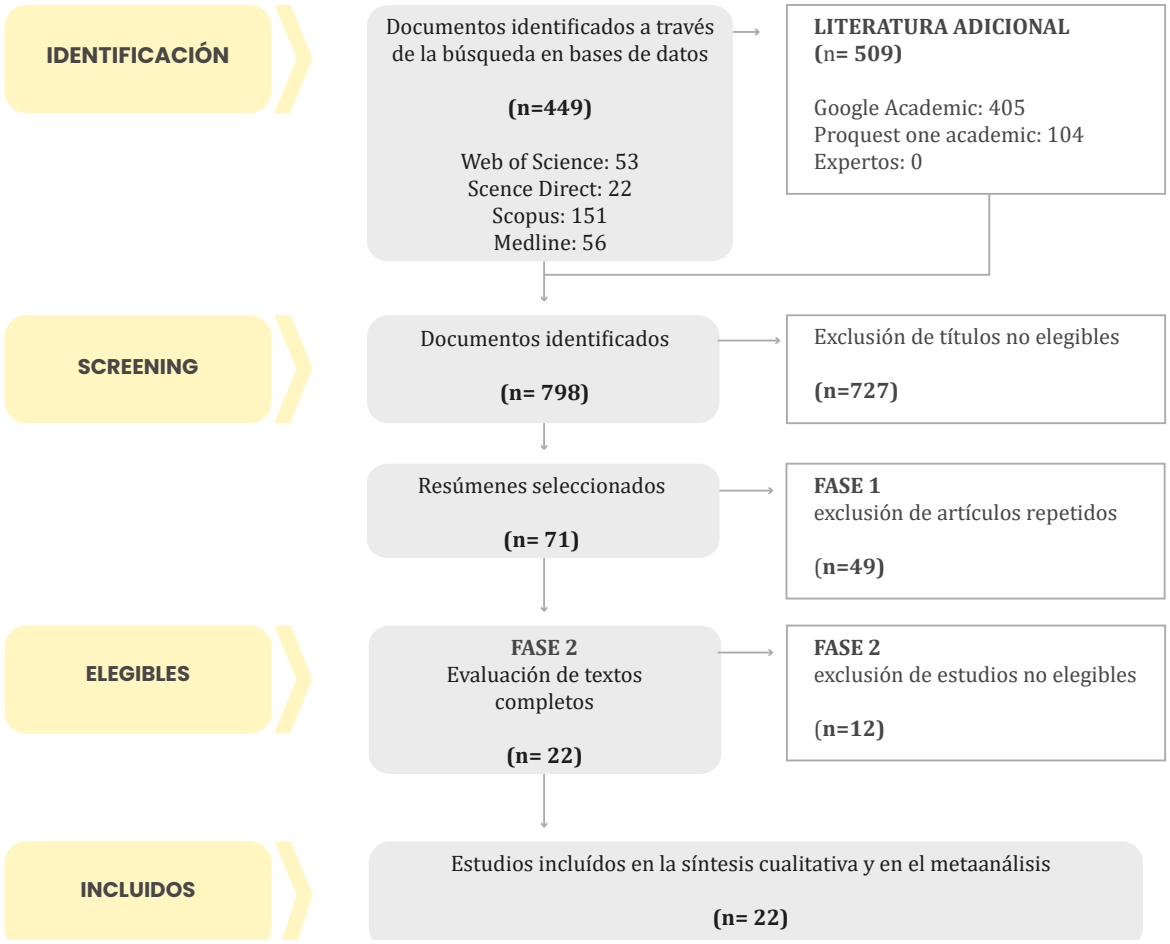


**Figura 1** Flujograma del proceso de selección

## CARACTERÍSTICAS DE LOS ESTUDIOS

Los 10 estudios incluidos [17,19,20,28–34] representaron un total de 4106 sujetos (2348 mujeres, 580 hombres, 878 género no especificado, con una proporción entre género masculino y femenino de 1 a 4). El rango de edad de estos estudios abarcó desde los 11 hasta los 78 años con una edad promedio de 3.7 años.

Los estudios se realizaron en 7 países diferentes con tamaños de muestra que variaron desde 102 hasta 923 personas. Las características descriptivas de estos estudios se resumen en la **tabla1**.

## RIESGO DE SESGO Y APLICABILIDAD DE LOS ESTUDIOS

Las diferencias metodológicas de los estudios incluidos obligaron a evaluar los aspectos que determinan la validez interna y externa de cada uno de ellos. Por medio de la herramienta QUADAS 2 [27] se pudo establecer el riesgo de sesgo de los estudios primarios (validez interna). En este sentido ninguno de los estudios cumplió con todos los criterios de calidad metodológica. Cinco de ellos [19, 31, 32, 34, 35] se consideraron con un alto riesgo de sesgo. Cuatro se consideraron de bajo riesgo [17,20,28,36] y uno se definió como no claro [35].

## TABLA 1

Resumen de las características principales de los estudios

| AUTORES/ LUGAR | Nº DE PARTICIPIANTES | | TIPO DE MUESTREO | ENTORNO | DISEÑO DEL ESTUDIO | PRUEBA ÍNDICE/ COMPRADOR | PRUEBA DE REFERENCIA |
|---|---|---|---|---|---|---|---|
| | Hombres | Mujeres | | | | | |
| Lovgren et al. Suecia | 126 | 323 | Consecutivos | Secundario (Centro de dolor orofacial) | Observacional tipo Caso - Control | Comparador 3Q-TMD | DC-TMD |
| Lovgren et al. Suecia | 300 | | Por conveniencia | Primario | Observacional tipo Caso - Control | Comparador 3Q-TMD | DC-TMD |
| Stasiak et al. Brasil | 265 | | Por conveniencia | Secundario | Observacional tipo Cohorte | Índice IAF | RDC-TMD |
| Kaynak et al. Turquia | 45 | 160 | Por conveniencia | Primario | Observacional tipo Caso - Control | Índice IAF | RDC-TMD |
| Zhang et al. China | 613 | | Por conveniencia | Primario | Observacional tipo Caso - Control | Índice IAF | DC-TMD |
| Dos Santos et al. Brasil | 0 | 203 | Por conveniencia | Primario | Observacional tipo Caso - Control | Índice IAF | DC-TMD |
| Ujin et al. China | 192 | 731 | Consecutivos | Secundario | Observacional tipo Caso - Control | Índice IAF | DC-TMD |
| Zagalaz - Anula et al. España | 25 | 77 | Por conveniencia | Secundario | Observacional tipo Caso - Control | Índice IAFC | DC-TMD |
| Ujin et al. China | 192 | 731 | Consecutivos | Secundario | Observacional tipo Caso - Control | Índice IAFC | DC-TMD |
| Pire et al. Brasil | 0 | 123 | Consecutivos | Secundario | Observacional tipo Caso - Control | Índice IAFC | RDC-TMD |

El mayor problema que presentaron tuvo relación con la selección de pacientes. Las recomendaciones de expertos en estudios de precisión diagnóstica (DTA) señalan que se debe evitar incluir pacientes seleccionados como casos y controles, dado que puede sesgar la precisión diagnóstica de una prueba. Como se ve reflejado en la tabla 1, solo 4 estudios incluyeron una muestra consecutiva de pacientes. En cuanto a la aplicabilidad, es decir, en qué medidas se pueden generalizar los resultados de estos estudios a otros pacientes y en otros lugares, todos los estudios presentaron un bajo riesgo de sesgo.

La tabla 2 representa una síntesis de los diferentes aspectos analizados en la evaluación del riesgo de sesgo.

## TABLA 2

Risk of bias assessment

| ESTUDIO | RIESGO DE SESGO | | | | PROBLEMA DE APLICABILIDAD | | |
|---|---|---|---|---|---|---|---|
| | Selección de pacientes | Prueba índice | Estándar de referencia | Flujo y sincronización | Selección de pacientes | Prueba índice | Estandar de referencia |
| Estudio 1 | 😊 | 😊 | 😊 | 😊 | 😊 | 😊 | 😊 |
| Estudio 2 | 😊 | 😊 | ? | 😊 | 😊 | 😊 | 😊 |
| Estudio 3 | 😊 | 😊 | 😠 | ? | 😊 | 😊 | 😊 |
| Estudio 4 | 😊 | 😊 | ? | 😊 | 😊 | 😊 | ? |
| Estudio 5 | ? | 😊 | ? | ? | 😊 | 😊 | ? |
| Estudio 6 | ? | 😠 | ? | 😊 | ? | 😊 | 😊 |
| Estudio 7 | 😠 | ? | 😊 | ? | 😊 | 😊 | 😊 |
| Estudio 8 | 😠 | ? | 😊 | 😠 | ? | 😊 | 😊 |
| Estudio 9 | 😊 | 😊 | 😊 | ? | 😊 | 😊 | 😊 |
| Estudio 10 | 😠 | ? | ? | ? | ? | 😊 | 😊 |

**Referencias:** 😊 Bajo riesgo 😠 Alto riesgo ? Riesgo no claro

## RESULTADOS DE ESTUDIOS INDIVIDUALES

A pesar de la gran heterogeneidad de los estudios se encontró una precisión diagnóstica muy buena tanto en la prueba 3Q-TMD como en el IAF. Para el diagnóstico de TTM se ha propuesto que los niveles aceptables de sensibilidad y especificidad sean de al menos 70% y 95% respectivamente [14].

A diferencia de los estudios diagnósticos, para los estudios de detección lo que interesaría es el nivel de sensibilidad a expensas de disminuir la especificidad dado que importaría que todos los individuos sospechosos sean detectados. Los dos estudios encontrados sobre 3Q-TMD utilizaron la prueba de referencia DC-TMD y obtuvieron los siguientes resultados de precisión diagnóstica: Para el estudio de Lovgren de la prueba 3Q-TMD en un entorno primario, los resultados de sensibilidad fueron de 0.81 (IC: 0.73-0.87) mientras que la especificidad fue de 0.79 (IC: 0.73-0.83). En un entorno secundario los valores de sensibilidad fueron de 0.96 (IC: 0.92-0.98) y los de especificidad 0.34 (IC: 0.28-0.40).

En cuanto a los estudios que examinaron la prueba IAF se encontró una gran heterogeneidad. Dividiremos los resultados de acuerdo a su versión original y a su

versión corta. Se encontraron cinco estudios de la versión original [19,31,32,35,36]. La mayoría muestran datos de sensibilidad y especificidad muy buenos.

El grupo de estudios de la versión corta del IAF se compone de 3 trabajos presentados entre 2018 y 2021. Los trabajos de Ujin-Yap y Pires [17,34] mostraron una sensibilidad y especificidad muy buena (0.95-0.93) para el estudio de Yap, y (0.86-0.95) para el de Pires. El estudio de Zagalaz Anula [37] mostró una sensibilidad y especificidad aceptable (0.78-0.79). En la tabla 3 se muestran los resultados de sensibilidad, especificidad y valores predictivos.

## TABLA 3

| Estudio | Año | Autor | Sensibilidad | IC 95% | Especificidad | IC 95% | VPP | VPN |
|---|---|---|---|---|---|---|---|---|
| Estudios 3Q-TMD | | | | | | | | |
| Estudio 1 | 2016 | Lövgren | 0.81 | 0.73-0.87 | 0.79 | 0.73-0.83 | 0.30 | 0.97 |
| Estudio 2 | 2018 | Lövgren | 0.96 | 0.92-0.98 | 0.34 | 0.28-0.40 | 0.59 | 0.90 |
| Estudios IAF | | | | | | | | |
| Estudio 3 | 2020 | Stasiak | 0.97 | 0.93-0.99 | 0.26 | 0.18-0.35 | 0.85 | 0.68 |
| Estudio 4 | 2020 | Kaynak | 0.94 | 0.86- 0.96 | 0.83 | 0.75-0.93 | 0.92 | 0.85 |
| Estudio 5 | 2019 | Zhang | 0.96 | 0.94- 0.97 | 0.72 | 0.59-0.83 | 0.97 | 0.62 |
| Estudio 6 | 2014 | Dos Santos | 0.86 | 0.78- 0.91 | 0.92 | 0.83-0.96 | 0.94 | 0.84 |
| Estudio 7 | 2021 | Yap | 0.95 | 0.93- 0.96 | 0.88 | 0.78-0.95 | 0.99 | 0.52 |
| Estudios IAF abreviado | | | | | | | | |
| Estudio 8 | 2021 | Zagalaz Anula | 0.78 | 0.64-0.88 | 0.79 | 0.65-0.89 | 0.78 | 0.79 |
| Estudio 9 | 2021 | Yap | 0.95 | 0.93-0.96 | 0.93 | 0.86-0.97 | 0.99 | 0.42 |
| Estudio 10 | 2018 | Pires | 0.86 | 0.75-0.9 | 0.95 | 0.87-0.99 | 0.94 | 0.99 |

## SÍNTESIS DE RESULTADOS

A pesar del escaso número de artículos, se encontró una alta heterogeneidad entre los trabajos analizados debido a la variabilidad de las características de la muestra, la diversidad metodológica y el riesgo de sesgo. El resultado del metaanálisis mostró valores de sensibilidad muy buenos para, IAF y 3Q-TMD: 0.92 (0.88-0.95). Por su parte los valores de especificidad mostraron una precisión cercana al 80%: 0.79 (0.63-0.90)

## METAANÁLISIS

**Sensibilidad:** En términos generales, todos los estudios mostraron una muy alta sensibilidad al usar un modelo de efectos fijos: (0.94) IC 95% (0.93-0.95). Cuando se observó el resultado bajo un modelo de efectos aleatorio, la sensibilidad para todas las pruebas de screening fue ligeramente más baja con un rango de fluctuación más amplio: (0.92) IC (0.88-0.95). Al discriminar entre las pruebas Índice (IAF) y comparador (3Q-TMD) no se hallaron diferencias significativas. (Figura 2)

**Especificidad:** En las pruebas de detección para TTM se observó un menor valor de especificidad que de sensibilidad y una mayor heterogeneidad entre los resultados. Usando un modelo de efectos aleatorios el metaanálisis mostró una especificidad de (0.79) IC 95% (0.63-0.90). (Figura 3)

**Medidas Adicionales:**

- **Odds Rato Diagnóstico (ORD):** Los valores globales de ORD para el conjunto de pruebas fueron de (3.81)

IC 95% (3.10-4.53). Observando la tabla con las estimaciones puntuales para el Odds Ratio.

Diagnóstico, se puede decir que los tests bajo análisis en cuanto a su precisión diagnóstica, muestran una alta sensibilidad y especificidad, lo que los muestran como buenas herramientas de detección. (**Figura 4**)

**Proportional hazards model approach:** Se utilizó el enfoque del modelo de riesgos proporcionales para estimar el comportamiento de las pruebas sobre una curva de rendimiento diagnóstico (ROC). Considerando este modelo bajo homogeneidad, los resultados globales de las pruebas IAF Y 3Q-TMD estimaron un área bajo la curva (AUC) de 0.95 IC 95% (0.97-0.92). Para el modelo de heterogeneidad los resultados fueron ligeramente menores manteniendo un desempeño muy bueno AUC 0.94 IC 95% (0.97-0.91). (**Figura 5**)

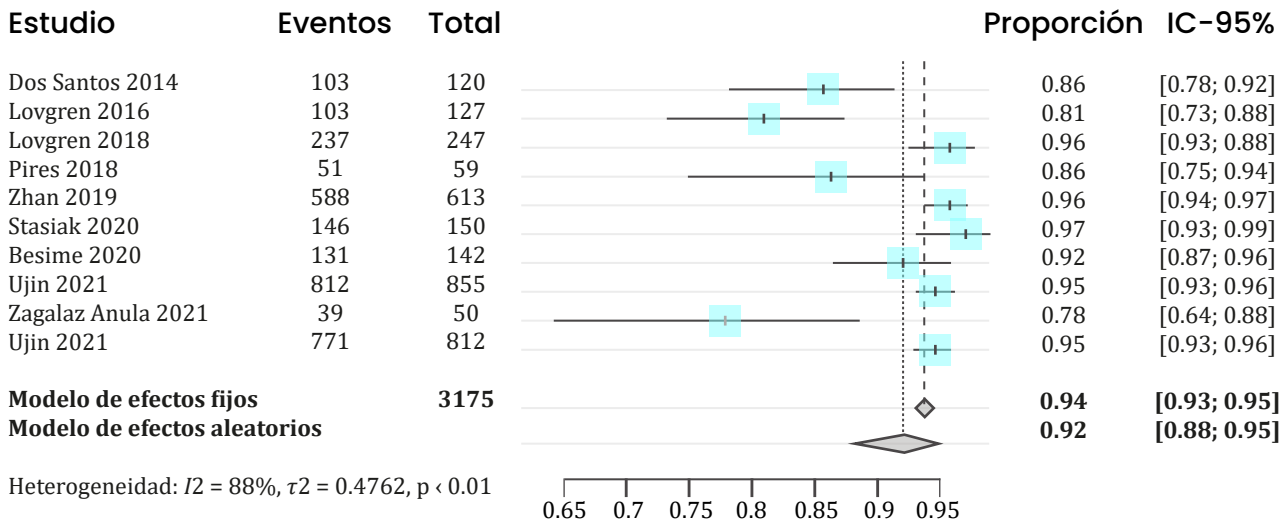


**Figura 2** Diagrama de bosque de la sensibilidad

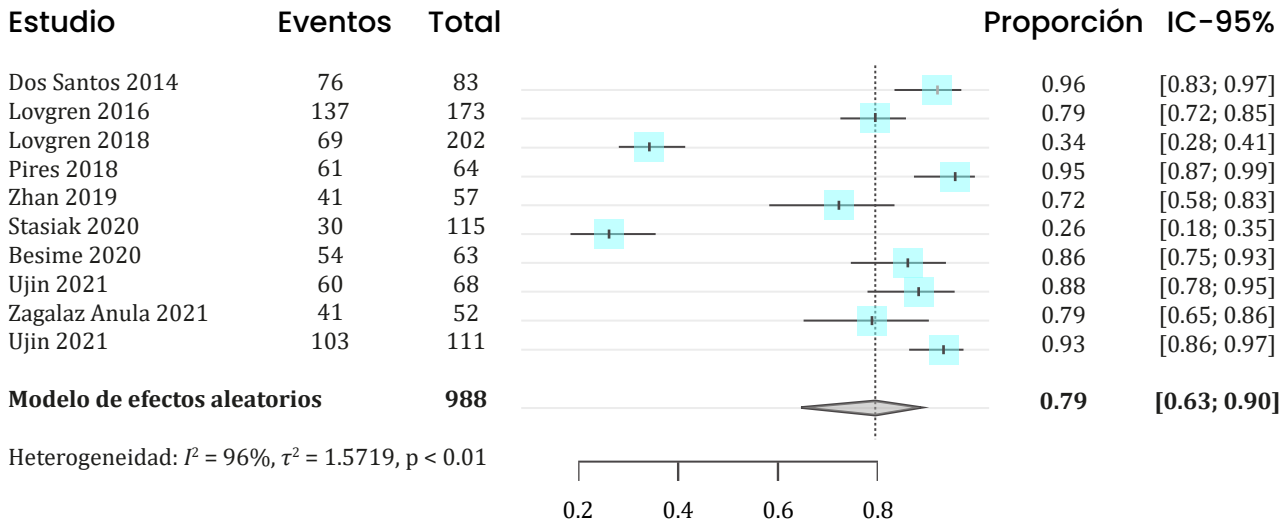


**Figura 3** Diagrama de bosque de la especificidad

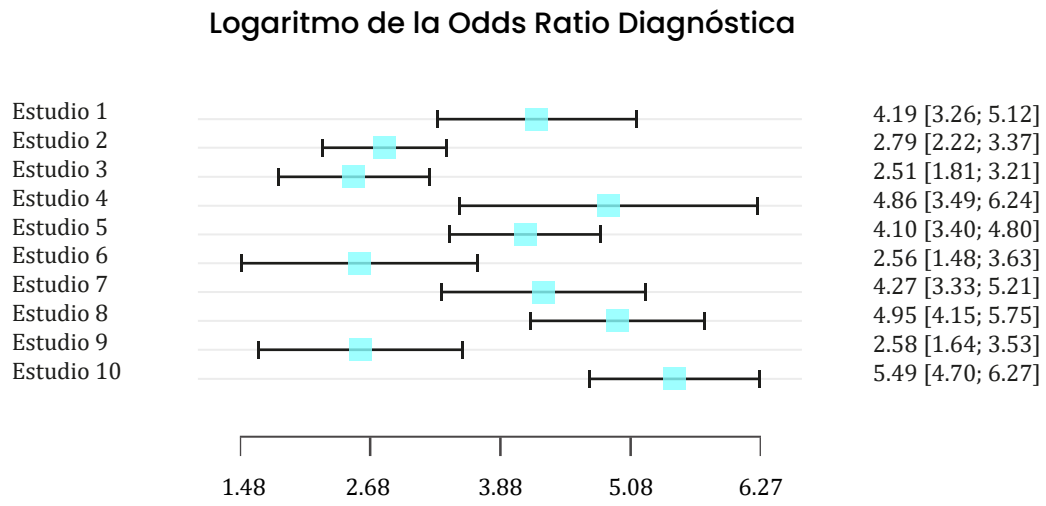


**Figura 4** Diagrama de bosque del Odds Ratio Diagnóstico

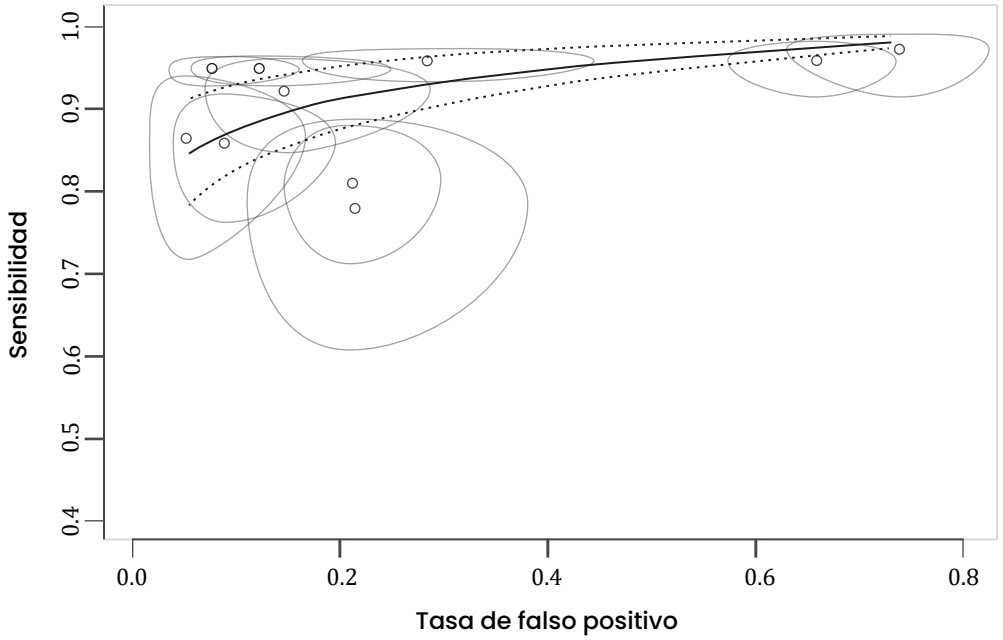


**Figura 5** Área bajo la Curva

# Discusión

El diagnóstico precoz es esencial en la atención médica, ya que define la condición y confirma el padecimiento del paciente. En algunas ocasiones, como en el caso de los TTM, establecer un diagnóstico es una tarea desafiante que puede ser facilitada por instrumentos de detección.

Este trabajo se basó en la comparación de dos pruebas de detección simplificadas a partir de su precisión frente a una prueba de diagnóstico de referencia (gold standard) con el fin de establecer la viabilidad de su aplicación como herramientas de cribado.

## PRUEBAS DE REFERENCIA

El desarrollo de RDC/TMD en 1992 tenía la intención de ser utilizado sólo para la investigación. Más tarde, en 2014, el DC/TMD amplió su uso al ámbito clínico. Estas herramien

tas diagnósticas están destinadas a establecer criterios confiables, estandarizados y validados para diagnosticar subtipos de TTM, ya que uno de los mayores problemas metodológicos en la investigación correlacionada es la definición precisa de los criterios aplicados [14]. La validez de criterio para las pruebas de detección se estableció en relación con el DC-TMD. Si bien el DC/TMD es fiable y válido, su uso rutinario para el triaje clínico de TTM no es práctico, dado que su protocolo de evaluación consume mucho tiempo y requiere la adecuada interpretación de sus complejos algoritmos [17]. Por su parte, las pruebas de screening para la detección de TTM permiten determinar de manera rápida y simple qué pacientes se beneficiarían de un diagnóstico específico.

Cabe señalar que la mayoría de los estudios de detección de esta revisión utilizaron el DC-TMD como prueba de referencia. Aunque el DC-TMD es la versión actualizada de los criterios del RDC- TMD, no se observó una diferencia sustancial en la precisión diagnóstica de los estudios cuando se los examinó a través del metaanálisis.

## VALIDEZ DE CRITERIO DE LA PRUEBA 3Q-TMD

La validez de criterio para la prueba 3 Q-TMD se estableció en relación al estándar de referencia DC-TMD en dos muestras de pacientes. Una de ellas sobre la población general y la otra sobre pacientes que acudían a un centro especializado. Cuando se compararon las respuestas positivas a las preguntas del 3Q-TMD en relación a la prueba de referencia se observó una diferencia sustancial entre los entornos. Sin embargo, la razón principal de esta diferencia se puede deber al cuestionario de síntomas del DC-TMD [16].

Al examinar los resultados en cada ámbito surgieron dos diferencias importantes. Primero, el marco temporal: si bien es real que ambos cuestionarios se basan en los síntomas informados, el test 3Q- TMD abarca un período semanal, mientras que la prueba de referencia se basa en los síntomas percibidos en los últimos 30 días. Segundo, la formulación de las preguntas: en el cuestionario 3Q-TMD la primera pregunta, se centra en el dolor reportado y la segunda en el dolor en la función. Sin embargo, para calificar para un diagnóstico de dolor DC/TMD, los criterios requieren que el dolor sea provocado o modificado por la función. Las diferencias observadas en la sensibilidad y especificidad del 3Q-TMD pueden obedecer a estas dos diferencias. Aun así, los valores predictivos fueron altos, especialmente los valores predictivos negativos, lo que indica que son preguntas excelentes para descartar un diagnóstico de TTM.

Un párrafo aparte merece la pregunta relacionada a los trastornos funcionales. Actualmente está en discusión el pronóstico de los sonidos articulares y las posibilidades de una evaluación clínica de un TTM intraarticular [38]. El DC/TMD ha mostrado una validez de moderada a pobre para un TMD intraarticular. En este sentido, el estándar de referencia actuaría mejor con fines de detección que como un instrumento diagnóstico [39]. Recientemente se ha puesto en duda la validez de los protocolos clínicos tanto del RDC-TMD como del DC-TMD para evaluar los trastornos intraarticulares [40]. Se proponen como alternativa de mayor precisión los estudios por imágenes como la resonancia, tomografía e incluso la artrografía. Es necesario subrayar que estos métodos estarían justificados sólo si el resultado de estos estudios puede cambiar el protocolo terapéutico. Sin embargo, es importante tener en cuenta que incluso en individuos asintomáticos, los desplazamientos de disco pueden estar presentes en la resonancia magnética en aproximadamente el 30% de la población [41,42]. A diferencia de las dos preguntas sobre dolor, los resultados de la pregunta sobre los trastornos intraarticulares mostraron una utilidad diferente según el entorno en el que fue aplicada. Cuando se aplicó sobre la población general fue útil para descartar la ausencia de un TTM intraarticular, mientras que, aplicada en un ámbito especializado su elevado valor predictivo positivo indicó un probable TTM intraarticular. En síntesis, aun cuando la validez de la pregunta sobre TTM intraarticular tiene una validez de regular a moderada su especificidad se mantuvo alta, lo cual indica que es muy útil a los fines de detección, particularmente para descartar un diagnóstico de disfunción cuando los resultados son negativos.

## VALIDEZ DE CRITERIO DE LA PRUEBA IAF

A diferencia del test 3Q-TMD, el índice Anamnésico de Fonseca identifica a los individuos con TTM a partir de un puntaje. Inicialmente el fin del IAF buscaba establecer el grado de severidad de los síntomas. Todos los estudios del IAF han mostrado altos niveles de precisión para detectar TTM relacionados con el dolor e intraarticulares. El presente estudio demostró que el área bajo la curva varió de entre 0.93 y 0.98 en diversas observaciones. [21,32,36]. Los estudios que analizaron el IAF establecieron un puntaje a partir de las preguntas que incluían los individuos con dolor, aquellos con trastornos intraarticulares y la suma de todos los síntomas. En este sentido, esta categorización fue similar a la analizada en el test 3Q-TMD. Las diferencias entre los estudios del IAF estuvieron en los puntos de corte. Cada estudio estableció puntos de corte diferentes para determinar la mayor precisión en relación a la prueba de referencia. La heterogeneidad de poblaciones y métodos podría explicar estas diferencias. Los puntos de corte del IAF para descartar individuos sin TTM fue de 0 a 20 puntos con un incremento del puntaje

de acuerdo a la intensidad de los síntomas. Si bien quedan dudas respecto a la utilidad de establecer un grado de severidad de los síntomas, los diferentes estudios han mostrado su capacidad de identificar TTM. En términos generales el IAF parece ser muy sensible para la detección, pero no muy específica. La menor especificidad observada con la IAF puede atribuirse a la inclusión de elementos no específicos de TTM como los dolores de cabeza, dolor de cuello, parafunción, maloclusión y tensión emocional. Esto derivó en una investigación de la dimensionalidad del cuestionario y sus propiedades psicométricas [43]. Este estudio confirmó la multidimensionalidad del IAF con una dimensión primaria de cinco ítems que dio como resultado el IAF corto de 5 preguntas. Los trabajos a partir de la versión abreviada del IAF mostraron un incremento de su precisión con un área bajo la curva de 0.97 y una mayor especificidad (95.5%) en relación a la prueba de referencia.

## ANÁLISIS DE LAS PRUEBAS IAF Y 3 Q-TMD

Los resultados del presente trabajo demostraron que, al comparar ambos test, el mayor número de estudios publicados sobre la prueba IAF permite considerar que los resultados observados en cuanto a su precisión diagnóstica son confiables, aún cuando se observó una gran heterogeneidad en los trabajos. Por su parte, sobre la prueba 3Q-TMD solo se encontraron 2 estudios. En este caso, la confiabilidad de sus resultados proviene de la evaluación de riesgo de sesgo, que para ambos estudios fue baja. Es importante señalar que, a pesar de la heterogeneidad de métodos y entornos, las pruebas de detección IAF y 3Q-TMD han mostrado muy buena precisión diagnóstica. Solo dos trabajos mostraron baja especi ficidad. El de Lovgren sobre las pruebas 3Q-TMD en un entorno especializado (especificidad 0.34) y el de Stasiak sobre el IAF (especificidad 0.26), también en un entorno secundario.

En cuanto al estudio de Lovgren que se realizó en una clínica especializada de dolor orofacial, la alta proporción de falsos positivos se explicaría porque en estos entornos la muestra de pacientes con dolor facial que no sea causado por TTM es mucho más frecuente que en un entorno primario. Es necesario remarcar que el DC/TMD fue elegido debido a su fiabilidad y validez para los diagnósticos de TTM más comunes. Sin embargo, hay una variedad de condiciones de dolor adicionales que pueden alterar la función normal de la mandíbula, como el dolor neuropático, la odontalgia atípica, la fibromialgia y el dolor cervical. Se espera que la prevalencia de afecciones dolorosas y no dolorosas sea mucho mayor en una clínica especializada en dolor orofacial en comparación con una clínica basada en la población general. También es lógico que la prevalencia de afecciones más raras sea mucho mayor en un entorno especializado [44]. Por lo tanto, las respuestas afirmativas al 3Q/TMD en una clínica especializada pueden estar relacionadas con un diagnóstico de TTM, pero también con varios diagnósticos diferenciales. En definitiva, esto podría explicar el aumento de falsos positivos con la consecuente baja en especificidad.

Se podría hipotetizar lo mismo sobre los resultados del estudio de Stasiak, sin embargo, hay motivos para pensar que esta diferencia podría deberse al riesgo de sesgo. Particularmente el estudio de Stasiak no define claramente cómo se aplicó el test de referencia ni cómo se estableció el flujo y la sincronización de los pacientes. El estándar de referencia utilizado para estos estudios comprende un sistema de diagnóstico basado en criterios estrictos y tanto la historia clínica como el examen están estructurados en detalle; los datos se procesan posteriormente en algoritmos predefinidos y dan como resultado un diagnóstico probable. Por otra parte, el examinador requiere un entrenamiento apropiado. Consecuentemente, un inadecuado manejo del test de referencia podría explicar esta discrepancia en los resultados.

En sintonía con esta explicación se encuentra el estudio de Yap, que utilizó el IAF y también fue realizado en un entorno secundario que, sin embargo, mostró una especificidad mucho mayor (0.88). En base a esta información no es posible establecer una explicación para las diferencias encontradas sobre la especificidad en entornos secundarios entre la prueba IAF. En conjunto, estos datos muestran la necesidad de más estudios de este tipo para explicar estas diferencias. Sin embargo, es necesario aclarar que al considerar estas pruebas como herramientas de detección y no de diagnóstico, la pérdida de especificidad no constituye una limitación. De hecho, es esperable que los test de detección o cribado no excluyan ningún potencial enfermo, aun cuando el "costo" a asumir sea un incremento de la presencia de falsos positivos. En este sentido, a diferencia de los estudios diagnósticos, para los estudios de detección lo que interesaría es el nivel de sensibilidad a expensas de disminuir la especificidad, dado que importaría que todos los individuos sospechosos sean detectados [45].

En cuanto a la sensibilidad, los resultados de casi todos los estudios mostraron una muy buena capacidad para detectar TTM. con excepción del estudio de Zagalaz que mostró una disminución leve. (Sensibilidad 0.78) Este resultado puede ser puesto en duda, dado que es el estudio con mayor riesgo de sesgo. A pesar de esto, tanto el IAF como el 3Q-TMD reflejaron buenos resultados para detectar individuos con TTM.

Aunque puedan existir dudas sobre aspectos puntuales de los resultados obtenidos, la falta de herramientas probadas para la detección de TTM señala que esta evidencia,

aunque no concluyente, es la mejor disponible a la actualidad e indica que se promueva el uso de estos instrumentos. Dado que el test 3Q-TMD ha sido validado en su lengua original y en el caso del IAF solo en 3 idiomas más, debería alentar a los investigadores a validar cualquiera de estos instrumentos en su propio idioma.

En esta revisión sistemática podrían señalarse algunas limitaciones. Por un lado, se encontraron pocos estudios, con resultados variables en cuanto a la precisión diagnóstica de cada prueba y una gran heterogeneidad metodológica. Por otro lado, se observó un alto riesgo de sesgo.

### DIRECCIONES FUTURAS

Como orientación hacia futuros estudios, recomendamos a los investigadores poner especial cuidado en la selección de pacientes. Particularmente, se debe evitar incluir pacientes bajo la forma “casos y controles” dado que puede sesgar los resultados, tal como lo expresan los especialistas en pruebas de precisión diagnóstica. Llamativamente, de todos los estudios analizados, este fue el aspecto que mayor influencia tuvo en el análisis de riesgo de sesgo. La heterogeneidad y el escaso número de los estudios encontrados, es una muestra de que las pruebas de detección de TTM, recién se están comenzando a considerar como herramientas necesarias en el ámbito del dolor craneofacial de origen no odontogénico. Esta observación es importante en esta revisión, dado que, aunque podamos tener dudas de los resultados obtenidos, la falta de herramientas probadas para la detección de TTM señala que esta evidencia, aunque no concluyente es la mejor disponible a la actualidad e indica que se promueva su uso. Como estos instrumentos han sido validados en su lengua original y en el caso del IAF solo en 3 idiomas más, se alienta a los investigadores a validar cualquiera de estos instrumentos en su propio idioma.

## Conclusiones

Las pruebas IAF Y 3Q-TMD son cuestionarios formulados con preguntas breves y constituyen herramientas simples y prácticas que permiten su aplicación de rutina en la clínica, sin afectar la actividad cotidiana.

Los resultados de la presente revisión sistemática y metaanálisis permiten concluir que tanto el IAF como EL 3Q-TMD son instrumentos muy sensibles para la detección de TTM en pacientes. Ambas pruebas de detección presentan la capacidad de identificar con mayor facilidad individuos con TTM. En este sentido, los pacientes detectados precozmente podrían beneficiarse de un diagnóstico y tratamiento oportunos, lo que evitaría largos periplos en busca de un diagnóstico y disminuir los costos de tratamientos, al evitar que estos problemas se cronifiquen.

Hasta donde sabemos, esta es la primera revisión sistemática con metaanálisis sobre pruebas de detección de TTM. La aparición de futuros estudios de validación permitirá incorporar nuevos datos y así obtener información más fidedigna sobre la capacidad de estas pruebas para detectar TTM.

## REFERENCIAS BIBLIOGRÁFICAS


**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.

## Disponibilidad de datos

Todo el conjunto de datos que apoya los resultados de este estudio fue publicado en el propio artículo.

## Declaración de conflicto de interés

Los autores declaran que no existen conflictos de intereses.

## Fuente de financiamiento

Los autores declaran no contar con fuente de financiamiento.

## Declaración de contribución de autoría y colaboración

| NOMBRE Y APELLIDO | COLABORACIÓN ACADÉMICA | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 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. Administración del proyecto
2. Adquisición de fondos
3. Análisis formal
4. Conceptualización
5. Curaduría de datos
6. Escritura - revisión y edición
7. Investigación
8. Metodología
9. Recursos
10. Redacción - borrador original
11. Software
12. Supervisión
13. Validación
14. Visualización

**Nota de aceptación:**
Este artículo fue aprobado por la editora de la revista MsC. Dra. Natalia Tancredi