



Comparación de dos pruebas diagnósticas para la detección de trastornos témporomandibulares. Una revisión sistemática y metaanálisis.

Autor: Hugo Daniel Rivera

*Maestría en Ciencias Odontológicas
Función Cráneo-mandibular y Dolor Orofacial
Facultad de Odontología
Universidad de la Republica*

2023



Comparación de dos pruebas diagnósticas para la detección de trastornos témporomandibulares. Una revisión sistemática y metaanálisis.

Autor: Hugo Daniel Rivera

Tesis presentada para la obtención del título de Magíster en Ciencias Odontológicas con énfasis en Función Cráneo-Mandibular y Dolor Orofacial, Facultad de Odontología – Udelar, Uruguay.

Tutor: Prof. Dr. Marcelo Kreiner, PhD

Agosto 2023

Dedicatorias

Esta tesis está dedicada a mi esposa, *María Angélica*, quién me animó y sostuvo en los momentos difíciles de este desafío y a mis hijas, *Camila* y *Paloma*, que me impulsan cada día a trabajar por un mundo mejor.

A la memoria **de mis padres** *Emilio Rivera* y *Rosa José* que seguro celebrarán con orgullo este logro.

Agradecimientos

Esta tesis no hubiera sido posible sin el apoyo de las siguientes personas, grupos humanos e instituciones:

Al Prof. Dr. PhD **Marcelo Kreiner**, como tutor y guía constante en este trabajo, gracias por la paciencia y la perseverancia en todas las idas y vueltas.

A La Dra. **Ivonne Rodriguez** de la Universidad Autónoma de México (UNAM) por haber dedicado sus conocimientos y su tiempo en la búsqueda y análisis de los trabajos que formaron parte de esta tesis. ¡Gracias por tu generosidad y excelente predisposición!

Al licenciado en estadística, Mag. **Ramón Álvarez** por facilitarnos la comprensión de las matemáticas que existen detrás de estos trabajos y por su disposición a colaborar en el desafío de esta tesis. Nuevamente ¡Muchas gracias!

A los compañeros de cursado de la Maestría y en particular a los que forman parte de la **Cátedra de Fisiología General y Bucodental** (Udelar), por su excelente recepción durante los momentos compartidos. ¡Siempre me sentí como en casa!

Al **Consejo de Facultad de Odontología** por el apoyo institucional que me permitió realizar este posgrado académico.

Al **Departamento de diagnóstico y tratamiento de los trastornos temporomandibulares de la Facultad de Odontología** (Udelar) y al Prof. Dr. **Raúl Riva**.

A mis colegas y amigos del Ateneo de Odontología de Rosario, **Bibiana, Silvana, Patricia, Ana** y **Mauro**, motores de esta locura de emprender una formación en el exterior.

A mi hermano **Gustavo**, compañero de algún viaje a Montevideo y cómplice incondicional de mis ilusiones.

A mi suegra **Marta** y mi suegro **Coco**, que desde donde esté, sentirá que su apoyo rindió sus frutos...

Tabla de contenidos

- 1. Resumen**
- 2. Abstract**
- 3. Abreviaturas**
- 4. Introducción**
 - 4.1 Trastornos temporomandibulares (TTM)
 - 4.2 Pruebas de diagnóstico para TTM
 - 4.3 Pruebas para la detección de TTM
 - 4.4 Precisión diagnóstica
 - 4.5 Planteamiento del problema y justificación
- 5. Objetivos**
 - 6.1 Objetivos generales
 - 6.2 Objetivos específicos
- 6. Hipótesis**
- 7. Material y método**
 - 7.1 Protocolo y registro
 - 7.2 Criterios de elegibilidad
 - 7.3 Identificación de las fuentes de estudio
 - 7.4 Estrategia de búsqueda
 - 7.5 Selección de estudios
 - 7.6 Proceso de recopilación de datos
 - 7.7 Riesgo de sesgo
 - 7.8 Medidas de precisión diagnóstica
 - 7.9 Síntesis de resultados
 - 7.10 Metaanálisis
- 8. Resultados**
- 9. Discusión**
- 10. Conclusiones**
- 11. Consideraciones éticas**
- 12. Referencias bibliográficas**

Resumen

Introducción

Los trastornos temporomandibulares (TTM) son afecciones frecuentes en la población general. En odontología son la principal causa del dolor orofacial después del dolor de origen dental. La prevalencia de los TTM se incrementa durante la adolescencia hasta llegar a su pico máximo en la edad media. 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. La falta de un diagnóstico oportuno causa incertidumbre y tiende a conducir a cuadros crónicos con consecuencias tanto económicas como en la calidad de vida del paciente.

Los métodos diagnósticos basados en criterios claramente definidos comenzaron a utilizarse a partir de 1992 con la aparición del protocolo RDC-TMD y su versión actualizada el DC-TMD. Desde entonces, estos protocolos han sido los métodos de diagnóstico de referencia tanto para la evaluación de los TTM en investigación como en la clínica. 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. Por esta razón, en los últimos años se comenzaron a validar métodos de detección más sencillos para identificar oportunamente a individuos que padecen TTM. La utilización rutinaria, por parte de clínicos generales, de pruebas de detección simplificadas de TTM permitiría un diagnóstico y tratamiento oportuno, lo que mejoraría la calidad de vida de los pacientes disminuyendo además los costos de tratamiento.

Objetivos

Analizar los métodos simplificados validados para detectar TTM. Específicamente, se ha comparado la precisión diagnóstica de las pruebas IAF y 3Q-TMD para la detección de TTM que tuvieran como gold estándar las pruebas RDC-TMD y DC-TMD.

Material y método

Se realizó una revisión sistemática con metaanálisis siguiendo las recomendaciones de las listas de verificación PRISMA-DTA. Se incluyeron estudios que evaluaron la precisión diagnóstica de los instrumentos IAF y 3Q-TMD a partir de las pruebas de referencia RDC y DC-TMD.

El período de búsqueda comprendió desde el año 1992 hasta 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.

Resultados

De 798 artículos, se incluyeron diez que representaron un total de 4106 sujetos. Los estudios se realizaron en 7 países diferentes con tamaños de muestra que variaron desde 102 hasta 923 personas. 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 IC (0.88:0.95) y una especificidad de 0.79 IC (0.63- 0.90) La medida de odds ratio diagnóstico que permite sintetizar en un solo valor la performance diagnóstica de una prueba mostró para el conjunto un resultado de (3.81) IC 95% (3.10; 4.53) lo cual indica que los tests bajo análisis se muestran como buenas herramientas de detección. Al analizar el comportamiento de las pruebas sobre una curva de rendimiento diagnóstico (ROC) se observó un desempeño muy bueno con un área bajo la curva de 0.94 IC 95% (0.97; 0.91).

Conclusión

A partir de este trabajo es posible concluir que tanto el IAF como EL 3Q-TMD son instrumentos sensibles para la detección de individuos con TTM. Ambas pruebas de detección tienen el potencial de permitir identificar con mayor facilidad individuos que padecen TTM y que podrían beneficiarse luego de un diagnóstico y tratamiento oportunos. De esta manera se podría mejorar la calidad de vida de los pacientes y disminuir los costos por tratamiento de afecciones crónicas.

Palabras clave: Trastornos Cráneo-mandibulares. Trastornos de la Articulación Temporomandibular. Triage. Dolor orofacial

Abstract

Introduction

Temporomandibular disorders (TMD) are common conditions in the general population. In dentistry they are the main cause of orofacial pain after dental pain. The prevalence of TMDs increases during adolescence and peaks in middle age. TMD-related pain can affect an individual's ability to perform daily activities, psychosocial functioning and

quality of life. The lack of timely diagnosis causes uncertainty and tends to lead to chronic conditions with consequences in both the economy and in quality of life the patient.

Diagnostic methods based on clearly defined criteria began to be used in 1992 with the appearance of the RDC-TMD protocol and its updated version DC-TMD. Since then, these protocols have been the gold standard diagnostic methods for both research and clinical evaluation of TMDs. While the DC-TMD is reliable and valid, its routine use for clinical triage of TMDs is not practical since its evaluation protocol is time-consuming and requires proper interpretation of its complex algorithms. For this reason, in recent years, simpler screening methods have begun to be validated to identify individuals with TMD in a timely manner. The routine use of simplified screening tests for TMD by general clinicians would allow timely diagnosis and treatment, which would improve the quality of life of patients and reduce treatment costs.

Objectives

To analyze the validated simplified methods to detect TMD. Specifically, we compared the diagnostic accuracy of the IAF and 3Q-TMD tests for the detection of TMD with the RDC-TMD and DC-TMD tests as the gold standard.

Material and Method

A systematic review with meta-analysis was performed following the recommendations of the PRISMA-DTA checklists. We included studies that evaluated the diagnostic accuracy of the IAF and 3Q-TMD instruments from the RDC and DC-TMD reference tests. The timeframe of the studies considered spanned from 1992 to April 2022 and 6 electronic databases were consulted. Two independent reviewers selected the studies. Risk of bias and applicability were assessed using the QUADAS 2 instrument.

Results

Of 798 articles, 10 representing a total of 4106 subjects were included. The studies were conducted in 7 different countries with sample sizes ranging from 102 to 923 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 CI (0.88:0.95) and a specificity of 0.79 CI (0.63- 0.90). The diagnostic odds ratio measure that allows synthesizing in a single value the diagnostic performance of a test showed for the set a result of (3.81) IC 95% (3.10; 4.53) which indicates that the tests under analysis are shown to be good screening tools. When analyzing the behavior of the tests on a diagnostic performance curve (ROC) a very good performance was observed with an area under the curve of 0.94 95% CI (0.97; 0.91).

Conclusion

From this work it is possible to conclude that both the IAF and EL 3Q-TMD are sensitive instruments for the detection of individuals with TMD. Both screening tests have the potential to more easily identify individuals with TMD who could benefit from timely diagnosis and treatment. This could improve the quality of life of patients and reduce costs for treatment of chronic conditions.

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

Abreviaturas:

ATM	Articulación temporomandibular
TTM	Trastornos temporomandibulares.
RDC-TMD	Criterios diagnósticos para la investigación de trastornos temporomandibulares
DC-TMD	Criterios diagnósticos para trastornos temporomandibulares
IAF	Índice anamnésico de Fonseca para la detección de TTM
IAFC	Índice anamnésico de Fonseca de formato corto
3Q-TMD	Test de 3 preguntas para la detección de TTM
PPD	Pruebas de precisión diagnóstica
VPP	Valores predictivos positivos
VPN	Valores predictivos negativos
LR +	Likelihood ratio +
LR-	Likelihood ratio -
ORD:	Odds ratio diagnóstico
ROC	Receiver operating characteristic

Introducción

El dolor y la disfunción temporomandibular afecta significativamente la calidad de vida de los pacientes (Willeman et al., 2014). Sin embargo, para el odontólogo general resulta complejo identificar a estos pacientes en sus etapas iniciales cuando los signos o síntomas pueden llegar a ser subclínicos. Es probable que la limitada formación de grado en esta área del conocimiento, sea un factor importante que contribuya a esta problemática. La detección precoz de estos trastornos daría a estos pacientes la posibilidad de un diagnóstico y tratamiento oportuno. Desde una perspectiva clínica, el odontólogo se beneficiaría al disponer de métodos simples y eficaces para detectar TTM en la práctica clínica. El hecho de contar con métodos de detección de TTM simples y eficaces podrían servir como una herramienta conveniente, tanto para detectar individuos con una posible necesidad de tratamiento, como también para aplicarla en estudios epidemiológicos.

Trastornos Temporomandibulares

La función del sistema masticatorio está estrechamente relacionada tanto con la supervivencia como con la interacción social. La importancia del sistema masticatorio también se expresa en la definición de salud bucal como "la salud oral es multifacética e incluye la capacidad de hablar, sonreír, oler, saborear, tocar, masticar, tragar y transmitir una variedad de emociones a través de expresiones faciales con confianza y sin dolor, incomodidad y enfermedad del complejo craneofacial"(fdi, 2020). Las afecciones más frecuentes del sistema masticatorio, después de la enfermedad dental, son los TTM (Lipton et al., 1993; Wirz et al., 2010).

Los TTM son afecciones que se caracterizan clínicamente por dolor y/o disfunción en los músculos masticatorios, cervicales, articulaciones temporomandibulares (ATM) y estructuras adyacentes (International Classification of Orofacial Pain, 1st Edition [ICOP], 2020).

En 1983 la asociación dental americana adoptó el término "trastornos temporomandibulares" con el objetivo de coordinar los esfuerzos de investigación con una terminología uniforme (Griffiths, 1983). Los síntomas indicativos de TTM pueden ser: dolor durante los movimientos de la mandíbula, alteración de la función mandibular y sonidos articulares (Okeson, 2019). Estos trastornos se pueden clasificar en dos subgrupos: los de origen articular, en los que los signos y síntomas están relacionados con la ATM, y los de origen muscular, cuando los signos y síntomas están relacionados con la

musculatura estomatognática. En particular, los TTM de origen articular abarcan varias alteraciones que afectan a los tejidos duros y blandos de la ATM. Entre los trastornos articulares más comunes se encuentran los trastornos del disco, dolor articular y enfermedad degenerativa (Schiffman et al., 2014). Por su parte, los TTM son la principal causa de dolor orofacial crónico (Lipton et al., 1993; Okeson, 2019; Wirz et al., 2010).

Dolor orofacial crónico

El dolor se define como "una experiencia sensorial y emocional desagradable asociada con, o parecida a la asociada con, daño tisular real o potencial"(Raja et al., 2020). La sensación de dolor agudo o protector puede causar reflejos de evitación y, por lo tanto, prevenir posibles daños tisulares. Cuando el dolor continúa durante un período más largo, e incluso después de que el tejido dañado se haya curado, se conoce como dolor no protector.

El grado de dolor y sufrimiento se relaciona más estrechamente con la amenaza que el paciente percibe de la lesión y la cantidad de atención que le presta, por este motivo se ha propuesto una revisión de la definición de dolor para subrayar el componente estresante y abarcar los aspectos sociales (Williams & Craig, 2016). La parte emocional de la sensación de dolor se describe como una experiencia multifactorial que incluye dimensiones sensoriales, emocionales, motivacionales y cognitivas. Estos aspectos están incluidos en el modelo biopsicosocial del dolor (Adler, 2009). El dolor crónico tiene un impacto negativo tanto en la vida social como laboral particularmente cuando es de carácter moderado a severo. Este tipo de dolor crónico afecta aproximadamente al 20% de la población adulta de Europa (Breivik et al., 2006). En Sudamérica se estima que el dolor crónico involucra alrededor del 30% de la población (Diaz, 2009). En odontología el dolor orofacial crónico está relacionado mayoritariamente a los TTM (Lipton et al., 1993; Wirz et al., 2010).

Epidemiología de los TTM

Los TTM son un problema de salud pública significativo que afecta aproximadamente al 30% de la población general y se considera la causa más común de dolor crónico de origen no dental en el área orofacial (Valesan et al., 2021) (List & Jensen, 2017). Los TTM pueden generar dolor y disfunción mandibular. Las condiciones más comunes de dolor son la mialgia, artralgia o cefalea asociada a TTM (Schiffman et al., 2014). Por su parte, los cuadros de disfunción están representados por dolor provocado por la función, ruidos y bloqueos

del movimiento mandibular, siendo la condición más frecuente de TTM en la población general la luxación del disco articular con o sin reducción (Manfredini et al., 2011). La luxación discal con y sin reducción son muy frecuentes en la adolescencia, el primero se asocia con ruidos durante la función mandibular mientras que la luxación del disco sin reducción se asocia con la restricción del movimiento mandibular especialmente en su etapa aguda. Por otra parte, los trastornos degenerativos de la articulación se presentan con mayor frecuencia en los adultos. Se manifiestan con dolor, crepitación articular y limitación de la función mandibular, aunque pueden no ser sintomáticos. Tanto los cuadros degenerativos como los trastornos del disco de la articulación suelen incorporar cuadros de mialgia (Schiffman et al., 2010).

La distribución de los TTM en las distintas etapas de la vida puede sintetizarse de la siguiente forma: la prevalencia se incrementa durante la adolescencia hasta llegar a su pico máximo en la edad media y se estima que se encuentran afectadas dos mujeres por cada hombre (List, 2005; Progiante et al., 2015). Los TTM dolorosos son más prevalentes que los trastornos intraarticulares no dolorosos (Machado et al., 2009; Manfredini et al., 2011). A su vez, sus signos y síntomas fluctúan con el tiempo (Marklund & Wänman, 2010; Wanman & Agerberg, 1986). Aun cuando la prevalencia y la distribución de los signos y síntomas indicativos de TTM están bien descritos sobre la base de estudios transversales, la etiología de estos trastornos aún se debate y no se desentraña por completo.

Etiología de los TTM

Por muchos años se consideró a la oclusión dental como causa principal de los TTM, sin embargo, al contrario de lo que se suponía, las alteraciones de la oclusión parecen ser más bien su consecuencia (Manfredini et al., 2017; Türp & Schindler, 2012). La evidencia actual indica que el origen de los síntomas de TTM es multifactorial, por esta razón se tienden a agrupar los numerosos factores en tres tipos posibles; los factores que predisponen la aparición de los síntomas, los que lo inician y los que los perpetúan (McNeill et al., 1980).

Entre los múltiples mecanismos predisponentes se encuentran los aspectos psicosociales como el catastrofismo, la depresión y condiciones genéticas (Melis & Di Giosia, 2016). Como factores iniciadores se han sugerido síntomas somáticos, bruxismo y ciertos trastornos del sueño (Fernandes et al., 2015; Fillingim et al., 2013; Finan et al., 2013; Sanders et al., 2013). Dentro de las condiciones que pueden perpetuar los síntomas, la somatización y la depresión son las más comunes (Epker & Gatchel, 2000; Ferrando et al., 2004).

Debido a todos estos hallazgos se ha adoptado un modelo de tipo biopsicosocial para la etiología de los TTM (Suvinen et al., 2005). Dado el cambio de modelo causal en relación al origen de estos trastornos, se ha identificado cierto grado de incertidumbre entre los odontólogos generales para el diagnóstico y tratamiento de los TTM (Durham et al., 2007).

Consecuencias de los TTM y necesidad de tratamiento

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 (Dahlstrm & Carlsson, 2010). De acuerdo a un metaanálisis realizado en 2008 la necesidad de tratamiento activo para los pacientes con TTM ronda entre el 1 y el 30% con una mediana del 16% de la población (Ameer Al-Jundi et al., 2008). La intensidad del dolor conduce al paciente a tomar múltiples citas y buscar diferentes modalidades terapéuticas (Van den Berghe et al., 2017). Se ha demostrado que para los individuos que sufren de TTM, la falta de un diagnóstico causa incertidumbre y tiene un impacto negativo en su vida diaria (Durham et al., 2010).

Pruebas de diagnóstico para TTM

Se han utilizado varios sistemas de diagnóstico para el dolor orofacial causado por TTM (Dworkin & LeResche, 1992; Schiffman et al., 2014). Uno de los más difundidos, el índice de Helkimo evaluaba paralelamente signos y síntomas con las condiciones oclusales (Helkimo, 1974). También la academia americana de dolor orofacial aportó un sistema de diagnóstico internacional, aunque sin criterios estrictamente definidos. En 1992 se establecieron los criterios de diagnóstico de investigación para TTM, conocido como RDC - TMD, cuyos fines era aplicarlo como instrumento de diagnóstico en investigación (Dworkin & LeResche, 1992). El RDC/TMD fue revisado exhaustivamente en los últimos años a partir de lo cual surgió su versión actual, el DC/TMD (Schiffman et al., 2014). La finalidad de este último es su aplicación no solo en el campo de la investigación, sino para el diagnóstico clínico. El DC/TMD se basa en extensos estudios clínicos multicéntricos, incluidos estudios financiados por el instituto nacional de salud de los EE. UU. y en conferencias internacionales de consenso.

Los protocolos diagnósticos RDC y DC -TMD han sido validados para las condiciones de TTM más comunes y han sido utilizados, en los últimos años, como prueba de referencia o gold estándar para evaluar otras pruebas de detección de TTM.

Esto obedece a que su uso rutinario para el cribado de pacientes con TTM no es práctico dado que su protocolo de evaluación consume mucho tiempo y requiere la adecuada interpretación de sus complejos algoritmos (Yap et al., 2021b).

Detección o Cribado de TTM

El cribado se define como el “examen médico efectuado a una persona para detectar una o varias enfermedades” (RAE, 2023). Un método de detección aceptable debe ser sencillo de aplicar y con costos equilibrados, pero también debe ser válido y en lo posible conducir a un tratamiento eficaz (Cochrane & Holland, 1971).

Pruebas para la detección de TTM

Durante los exámenes dentales de rutina es necesario identificar aquellos individuos que se beneficiarían de un examen completo para evaluar un posible TTM. En este sentido, los métodos de detección prometen ser precisos, fiables y económicos, pero además implican otros beneficios. Se ha sugerido que la detección temprana permitiría prevenir los cuadros crónicos (Macfarlane, 2016). A su vez, la mayoría de los pacientes con diagnóstico de TTM se beneficiarían con tratamientos conservadores y únicos (Magnusson et al., 2002).

Se han validado varios métodos diagnósticos para la detección de TTM. Los más estudiados y validados han sido el screening de dolor por TTM de la prueba RDC-TMD, el IAF (Stasiak et al., 2020) y el Test 3Q-TMD (Lovgren et al., 2016). De estas 3 pruebas, el screening RDC-TMD ha demostrado validez para detectar individuos con dolor potencial por TTM (Gonzalez et al., 2011). Sin embargo, este último no incluye aspectos funcionales no dolorosos (Lövgren, 2017).

Índice Anamnésico de Fonseca

Para realizar el cribado de los pacientes, en 1994 se propuso el IAF (Tabla 1), que evalúa la gravedad de los TTM en función de sus signos y síntomas. El IAF se basa en el índice de Helkimo (Helkimo, 1974). En comparación con este tiene una buena correlación positiva y confiabilidad. 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 (Berni et al., 2015; Pastore et al., 2018).

El IAF consta de 10 preguntas cuyo objetivo es obtener un índice a partir de la anamnesis del paciente. Cada pregunta tiene 3 opciones de respuesta con valores asignados: "Sí" (10

puntos), "A veces" (5 puntos) y "No" (0 puntos). Después de la suma de las respuestas, los pacientes se clasifican según la tabla de valores: de 0 a 15 (sin TTM); 20–40 (TTM leve); 45–65 (TTM moderada) y 70–100 (TTM severa).

IAF	IAF C	INDICE ANAMNESICO DE FONSECA	Si	A	No
1*	1	¿Tiene dificultad para abrir la boca de par en par?			
2*	2	¿Tienes dificultad para mover la mandíbula hacia los lados?			
3*	3	¿Sientes fatiga o dolor muscular cuando masticas?			
4*	--	¿Tiene dolores de cabeza frecuentes?			
5*		¿Tiene dolor de cuello o rigidez en el cuello?			
6*	4	¿Tiene dolores de oído o dolor en esa área (articulación temporomandibular)			
7*	5	¿Alguna vez has notado algún ruido en tu articulación temporomandibular mientras masticas o abres la boca?			
8*	--	¿Tienes algún hábito como apretar o rechinar los dientes?			
9*	--	¿Sientes que tus dientes no se juntan bien?			
10*	--	¿Te consideras una persona tensa (nerviosa)?			
		RESULTADO			
-- Las preguntas 4*, 5*, 8*, 9*, y 10* están excluidas en el IAF corto					

Tabla 1. Preguntas del IAF original (primera Columna) y del IAF Corto (segunda columna)

Recientemente fueron investigadas la dimensionalidad y las propiedades psicométricas de la IAF donde se observó multidimensionalidad con una dimensión primaria formada por 5 ítems fiables, concretamente las preguntas 1, 2, 3, 6 y 7, con un ajuste adecuado al modelo de Rasch para los trastornos musculares definidos por RDC/TMD (Rodrigues-Bigaton et al., 2017). A partir de ese trabajo se originó el IAF de formato corto que ofrece una excelente fiabilidad para los 5 elementos específicos de TTM, así como para todo el instrumento (Tabla 1). El IAF C además de ser más corto y rápido de administrar puede limitar la sobreestimación de la prevalencia o la presencia de TTM al excluir los ítems y factores de riesgo no específicos de los TTM (Bevilaqua-Grossi et al., 2014).

Test 3Q TMD

Originalmente este test surgió en Suecia con el objetivo de mejorar la calidad de la atención dental de los pacientes con TTM (Lovgren et al., 2016). Su implementación fue

obligatoria en el sistema de registro dental de una comunidad con el fin de identificar los pacientes con TTM más severo durante los chequeos dentales regulares.

El test 3Q-TMD (Tabla 2) se compone de un cuestionario de 3 preguntas, dos de las cuales se orientan hacia el dolor facial y la tercera se enfoca en los aspectos funcionales frecuentes de los TTM intraarticulares. Cada pregunta debe responderse con un "sí" o "no". El lapso de tiempo establecido como “una vez a la semana” es el que tiene mayor correspondencia con la necesidad de tratamiento percibida por los pacientes. La fiabilidad y validez de estas preguntas se han determinado en relación con el RDC/TMD en adolescentes (Nilsson et al., 2006).

	Test 3Q TMD	Si	No
P 1	¿Tiene dolor en la sien, la cara, la mandíbula o la articulación de la mandíbula una vez a la semana o más?		
P 2	¿Tiene dolor una vez a la semana o más cuando abre la boca o mastica?		
P 3	¿Su mandíbula se bloquea o se atasca una vez a la semana o más?		

Tabla 2. Preguntas del Test 3Q TMD

Los pacientes con TTM parecen no ser detectados y tratados oportunamente por la odontología (Nasstrom et al., 2019). Por esta razón, consideramos que es importante examinar los métodos de detección que permitirían identificar los TTM dolorosos y funcionales más comunes.

En este sentido los únicos dos tests de detección validados que abarcan estos aspectos son el 3Q- TMD y el IAF, por lo tanto, hemos centrado este trabajo en comparar su precisión diagnóstica.

Precisión Diagnóstica

Las pruebas diagnósticas se evalúan por medio de dos condiciones: la validez y la fiabilidad de la prueba en cuestión. La fiabilidad consiste en la capacidad para reproducir los mismos resultados cada vez que se aplica, mientras que la validez refiere a la precisión con la cual la prueba es capaz de detectar individuos con la afección buscada.

Existen diferentes tipos de validez. Una forma básica de validez es la "validez aparente", que se refiere a una evaluación subjetiva de si la prueba mide lo que se supone que debe medir. La validez del criterio, por otra parte, evalúa la capacidad de una prueba o método para predecir el resultado de otra prueba. La validez del criterio se utiliza comúnmente

para determinar la precisión diagnóstica de una prueba, como un método de detección, en comparación con una prueba de referencia.

Para evaluar la precisión de una prueba diagnóstica existen varios cálculos matemáticos con ventajas específicas, disponibles. La precisión diagnóstica de una prueba se puede informar como sensibilidad: que es la capacidad de una prueba índice para definir correctamente un resultado positivo de la prueba cuando la enfermedad está presente, y especificidad: la capacidad de una prueba para definir un resultado negativo de la prueba cuando la enfermedad está ausente (Eusebi, 2013). La sensibilidad y la especificidad de una prueba se calculan dentro de una "tabla de dos por dos" y se pueden resumir como la capacidad de una prueba para identificar "verdaderos positivos" y "verdaderos negativos". Estas estimaciones no tienen en cuenta la prevalencia o la probabilidad previa a la prueba. En la práctica clínica y de forma individual, el proceso de decisión clínica se basará en las quejas reales del paciente junto con los hallazgos en el examen. Por lo tanto, los valores predictivos de una prueba se pueden calcular para dar una estimación de la precisión para el paciente individual. En los cálculos de valores predictivos, se tiene en cuenta la prevalencia de la afección y, por lo tanto, afectará el resultado. El valor predictivo positivo es la probabilidad de que un individuo tenga una condición dado un resultado positivo de la prueba. El valor predictivo negativo es la probabilidad de que un individuo con un resultado negativo de la prueba no tenga la condición (Eusebi, 2013).

Cuando se estudia la precisión diagnóstica de una prueba es importante que se realice en un entorno representativo en relación con la situación clínica real (Aleem et al., 2009). Dentro de la práctica general, con una prevalencia relativamente baja de TTM junto con el hecho de que TMD es una condición no mortal, una prueba de diagnóstico para TTM debe tener preferiblemente una alta especificidad (Oakley & Brunette, 2002). Los objetivos sugeridos de precisión diagnóstica para una prueba diagnóstica de TTM han sido una sensibilidad >70% y una especificidad >95% (Dworkin & LeResche, 1992). Esto permitirá la identificación correcta de aquellos sin la condición y resultará en pocos tratamientos iniciados innecesarios.

Planteamiento del problema y justificación

Aun cuando los TTM tienen un impacto negativo en la vida de los pacientes, la atención dental de rutina parece no detectarlos adecuadamente. Esto se ve reflejado por la discrepancia entre la necesidad estimada de tratamiento y el tratamiento efectivamente

realizado (Ameer Al-Jundi et al., 2008; Wänman & Wigren, 1995). Solo cuando el dolor se hace crónico los pacientes tienden a buscar tratamiento (Lundh et al., 1987; Magnusson et al., 2000). 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 (Alex White et al., 2001). Se ha calculado que solo en EE. UU. el gasto por tratamiento de estas afecciones en la última década se ha duplicado, estimándose en unos 4.000 millones de dólares anuales (National Institute of Dental and Craniofacial Research. Facial Pain (NIDCR), 2022). 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 (Durham et al., 2011; Macfarlane, 2016; Stowell et al., 2007).

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. Por lo tanto, esta revisión sistemática se ha enfocado en responder la siguiente pregunta: ¿Cuál es la validez (especificidad y sensibilidad: precisión) de los instrumentos IAF y 3Q-TMD para la detección de TTM?

Objetivos

General: El objetivo de este trabajo fue evaluar y comparar la precisión diagnóstica de las pruebas IAF y 3Q-TMD para la detección de TTM.

Objetivos específicos:

- Analizar la precisión diagnóstica de cada uno de los instrumentos evaluados.
- Evaluar el riesgo de sesgo de los estudios publicados.
- Evaluar la heterogeneidad de los resultados de los estudios evaluados a partir de la selección de pacientes, las características del diseño de los estudios y los cambios en el umbral diagnóstico.
- Detectar posibles debilidades en estos instrumentos que puedan generar nuevas preguntas de investigación para el futuro avance de esta área del conocimiento.

Material y método

Protocolo y registro

El protocolo de esta revisión fue documentado de acuerdo con la guía de la declaración PRISMA-P (Shamseer et al., 2015) y fue inscrito en PROSPERO (International prospective register of systematic reviews). *Es necesario aclarar que el protocolo fue aceptado en PROSPERO, pero debido a la sobrecarga de estudios sobre COVID de ese momento, lo cual saturó el sistema, no se obtuvo número de registro.*

Para elaborar esta revisión sistemática de pruebas diagnósticas se siguieron las recomendaciones de la lista de verificación PRISMA-DTA (Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies) (Salameh et al., 2020).

Criterios de elegibilidad

En esta revisión se incluyeron estudios que evaluaron la precisión diagnóstica de los instrumentos IAF y 3Q-TMD a partir de las pruebas de referencia 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 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 de entornos primarios (centros odontológicos generales) y secundarios (centros especializados en dolor orofacial). Para establecer la condición objetivo, los pacientes de estos estudios respondieron a una evaluación clínica individual de TTM (particularmente ruidos articulares, limitación del movimiento mandibular, dolor muscular, articular y o dolor pre auricular) de acuerdo a los criterios establecidos por el consorcio INFORM (International Network for Orofacial Pain and Related Disorders Methodology [INFORM], 2023).

Los criterios de exclusión abarcaron lo siguiente: 1) estudios que no utilizaron RDC/TMD (estudios publicados antes de 1992) o DC/TMD, estudios que modificaron la herramienta; 2) artículos con datos duplicados de otro estudio incluido; 3) estudios enfocados exclusivamente en pacientes con prótesis completa, tratamiento de ortodoncia; 4) estudios centrados en muestras de pacientes con comorbilidades (fibromialgia, hipermovilidad articular sistémica, artritis idiopática juvenil, tinnitus, trastorno de estrés postraumático,

osteoartritis sistémica, neuralgia del trigémino, síndrome de boca ardiente, dolor facial atípico, migraña, odontalgia atípica, dolor cervical, dolor neuropático y deformidades dento-esqueléticas) o fracturas y trastornos congénitos/del desarrollo (aplasia, hipoplasia, hiperplasia) o cualquier otro síndrome asociado a TTM; 5) revisiones, cartas, libros, opinión de expertos e informes de casos.

Identificación de las fuentes de estudio

Para la estrategia de búsqueda se utilizó un filtro de búsqueda para estudios de precisión diagnóstica (Health Information Research Unit [HIRU]2022). 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 realizó una búsqueda de literatura gris en 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, Países Bajos).

Estrategia de Búsqueda

Se llevó a cabo una estrategia de búsqueda en las bases de datos previamente señaladas. Se utilizaron términos indexados y términos libres para localizar la investigación realizada sobre TTM, así como su precisión diagnóstica. Se describe a continuación la estrategia de búsqueda utilizando el filtro para estudios de precisión diagnóstica (HIRU, 2022):

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 (sensitivity*[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 revisores (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 (CIR) para asegurar la integridad de los contenidos. Cualquier desacuerdo se discutió 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. Cegamiento y forma de randomización de los pacientes para las pruebas. Entorno sanitario donde se realizó el estudio (atención primaria o secundaria).

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

En las pruebas IAF y 3Q-TMD, las definiciones de verdadero positivo se consideraron cuando los individuos respondían afirmativamente una de las preguntas relacionadas al dolor facial, función mandibular o limitación del movimiento mandibular y cuyos síntomas eran corroborados por la prueba de referencia. Verdadero negativo se consideró a aquellos individuos donde las preguntas de la prueba IAF y 3 Q-TMD eran respondidas negativamente y corroboradas por la prueba RDC o DC-TMD. La condición de falso positivo se estableció cuando las preguntas a las pruebas 3Q-TMD o IAF eran respondidas

positivamente pero el examen de referencia reflejaba ausencia de TTM. Los casos falsos negativos por su parte se identificaron cuando las preguntas de las pruebas IAF y 3Q-TMD se respondían negativamente, pero la prueba de referencia diagnosticaba algún tipo de TTM.

Riesgo de sesgo

El riesgo de sesgo fue evaluado de forma independiente por dos revisores (HDR y CIR). Se utilizó la herramienta QUADAS 2 para examinar el riesgo de sesgo y la aplicabilidad de los estudios (Ciapponi, 2015; Whiting et al., 2011). 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 adicionales como el odds ratio diagnóstico (ORD) y se utilizó el 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 (IAF y 3Q-TMD) en diagramas de bosque acoplados y en una curva de rendimiento diagnóstico (Curvas ROC).

Metaanálisis

Los metaanálisis de revisiones sistemáticas sobre precisión de pruebas diagnósticas (PPD) tienen un resultado bivariado, al considerar en general un par de métricas como son la sensibilidad y especificidad que fluctúan en forma simultánea. En particular toma especial relevancia el manejo de la heterogeneidad en PPD, que suele ser mucho mayor debido a la variación entre los estudios, por la forma en que un marcador continuo se dicotomiza a través de diferentes entornos provocando una variación en la precisión de las pruebas.

Debido a que no todos los estudios reportaban valores de prevalencia, fue necesario realizar su estimación. Para poder iniciar el metaanálisis se calcularon los valores de prevalencia a partir de los valores predictivos positivos (VPP) los valores predictivos negativos (VPN) y el tamaño muestral de cada estudio. Obtenida la prevalencia de cada estudio se estimaron los puntos operativos resumidos (sensibilidades y especificidades resumidas) tanto para la prueba 3Q- TMD como el IAF, bajo un intervalo de confianza del 95 %. A continuación, se estableció el ORD para obtener en un solo número la PPD 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 registros para la evaluación de títulos y resúmenes se eliminaron 49 títulos 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 para realizar la revisión sistemática y el metanálisis. Una visión general del proceso de selección se muestra en el gráfico 1.

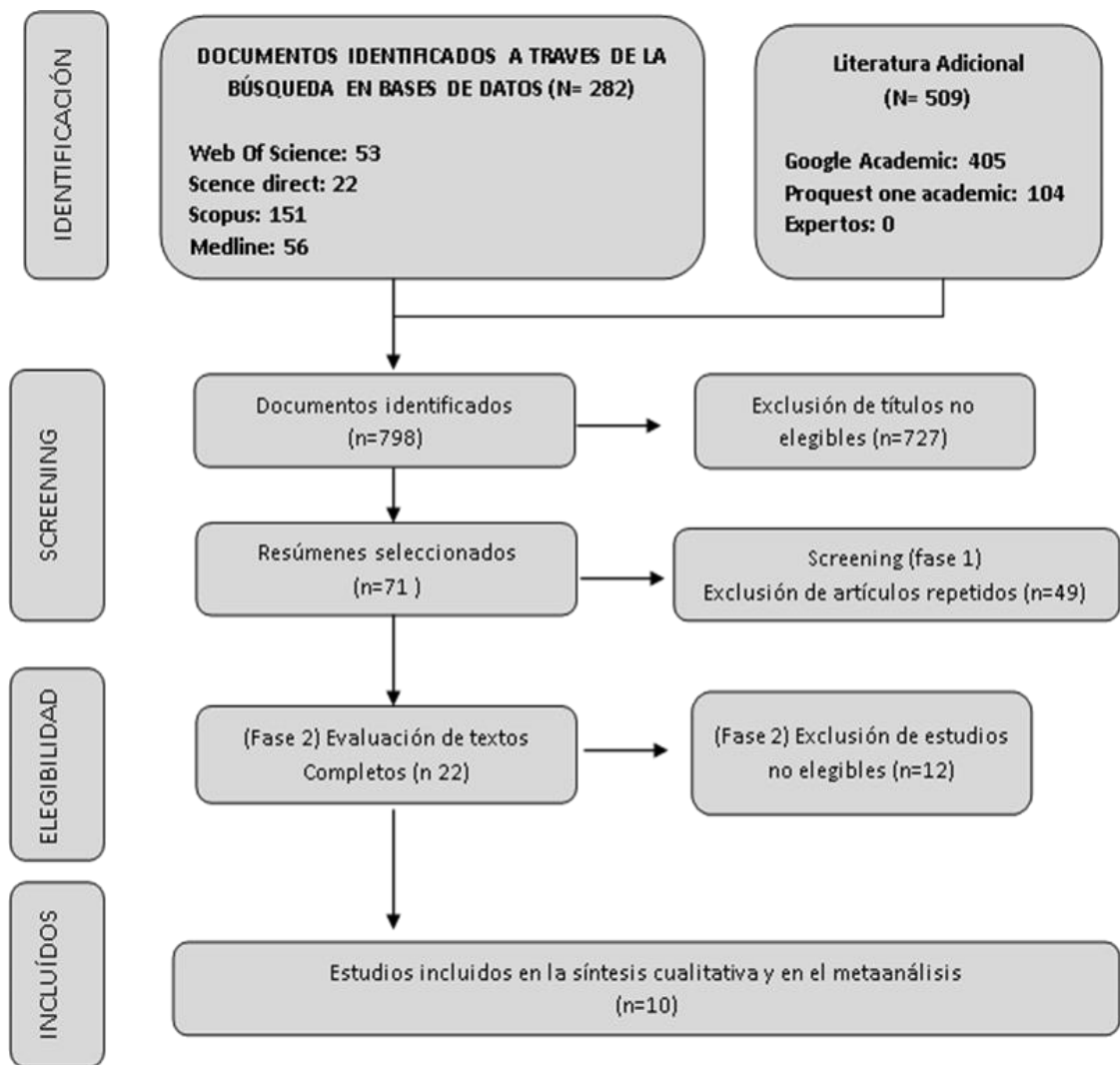


Gráfico 1. Flujograma del proceso de selección

Características de los estudios

Los 10 estudios incluidos representaron un total de 4106 sujetos (Berni et al., 2015; Kaynak et al., 2020; Lovgren et al., 2016, 2018; Pires et al., 2018; Stasiak et al., 2020; Yap et al., 2021a, 2021b; Zagalaz-Anula et al., 2021; Zhang et al., 2020a). Estaban compuestos por 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 33.7 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 tabla 3.

Autor	N° de participantes		Tipo de muestreo	Entorno	Diseño del estudio	Pruebas a Comparar	Prueba de referencia
	Hombres	Mujeres					
Lovgren	126	323	Consecutivos	Secundario (Centro de dolor orofacial)	Observacional tipo Caso - Control	3Q-TMD	DC-TMD
Lovgren	300		Por conveniencia	Primario	Observacional tipo Caso - Control	3Q-TMD	DC-TMD
Stasiak	265		Por conveniencia	Secundario	Observacional tipo Cohorte	Índice FAI	RDC-TMD
Kaynak	45	160	Por conveniencia	Primario	Observacional tipo Caso - Control	Índice FAI	RDC-TMD
Zhang	613		Por conveniencia	Primario	Observacional tipo Caso - Control	Índice FAI	DC-TMD
Dos Santos	0	203	Por conveniencia	Primario	Observacional tipo Caso - Control	Índice FAI	RDC-TMD
Yap	192	731	Consecutivos	Secundario	Observacional tipo Caso - Control	Índice FAI	DC-TMD
Zagalaz	25	77	Por conveniencia	Secundario	Observacional tipo Caso - Control	Índice SFAI	DC-TMD
Yap	192	731	Consecutivos	Secundario	Observacional tipo Caso - Control	Índice SFAI	DC-TMD
Pires	0	123	Consecutivos	Secundario	Observacional tipo Caso - Control	Índice SFAI	RDC-TMD

Tabla 3. Resumen de las características principales de los estudios

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 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 se consideraron con un alto riesgo de sesgo (Berni et al., 2015; Pires et al., 2018; Stasiak et al., 2020; Yap et al., 2021a; Zhang et al., 2020). Cuatro se consideraron de bajo riesgo (Kaynak et al., 2020; Lovgren et al., 2016, 2018; Yap et al., 2021b). Solo uno se definió como no claro (Zhang et al., 2020).

El mayor problema que presentaron tuvo relación con la selección de pacientes. Las recomendaciones de expertos en el estudio de pruebas de precisión diagnóstica señalan que se debe evitar la selección de pacientes como casos y controles dado puede sesgar la precisión diagnóstica de una prueba (Bossuyt & Leeflang, 2008). Como se ve reflejado en la tabla 4, solamente 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. La tabla 4 representa una síntesis de los diferentes aspectos analizados en la evaluación del riesgo de sesgo. Mayor información se presenta en los documentos anexos.

ESTUDIO	RIESGO DE SESGO				PROBLEMAS DE APLICABILIDAD		
	Selección de pacientes	Pruebas a comparar	Estándar de referencia	Flujo y sincronización	Selección de pacientes	Pruebas a comparar	Estándar de referencia
Lövgren 2016	😊	😊	😊	😊	😊	😊	😊
Lövgren 2018	😊	😊 ^e	?	😊	😊	😊	😊
Stasiak 2020	😊	😊	😞	?	😊	😊	😊
Kaynak 2020	😊	😊	?	😊	😊	😊	?
Zhang 2021	?	😊	?	?	😊	😊	?
Dos Santos 2014	?	😞	?	😊	?	😊	😊
Yap 2021	😞	?	😊	?	😊	😊	😊
Zagalaz 2021	😞	?	😊	😞	?	😊	😊
Yap 2021	😊	😊	😊	?	😊	😊	😊
Pires 2018	😞	?	?	?	?	😊	😊

Tabla 4. Resultados de QUADAS-2 Referencias: 😊 Riesgo Bajo 😞 Alto ? 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 (Dworkin & LeResche, 1992).

Los dos estudios encontrados sobre 3Q-TMD utilizaron la prueba de referencia DC-TMD y obtuvieron los siguientes resultados de precisión diagnóstica: el estudio de Lovgren de la prueba 3Q-TMD en un entorno primario, mostró resultados de sensibilidad de 0,81 (IC-0,73-0,87) mientras que la especificidad fue de 0,79 (IC-0,73-0,83). En cuanto al estudio de Lovgren que se realizó en una clínica especializada de dolor orofacial (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). Ambos estudios sobre 3Q-TMD muestran una muy buena sensibilidad, es decir una gran capacidad para detectar individuos con TTM y una buena especificidad solo cuando se utiliza en entornos de atención primaria. En cuanto a la capacidad para detectar a los individuos libres de TTM en un entorno especializado, la

prueba 3Q-TMD mostró resultados pobres. Por su parte, en 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 (Berni et al., 2015; Kaynak et al., 2020; Stasiak et al., 2020; Yap et al., 2021a; Zhang et al., 2020). La mayoría muestran datos de sensibilidad y especificidad muy buenos, excepto el estudio de Stasiak (Stasiak et al., 2020), que al igual que el estudio de Lovgren, tuvo una sensibilidad muy alta (0.97) pero una especificidad muy baja (0.26).

El grupo de estudios de la versión corta del IAF se compone de 3 trabajos presentados entre 2018 y 2021. Los trabajos de Yap y Pires (Pires et al., 2018; Yap et al., 2021b) 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. En este grupo como en los anteriores también se encontró una discrepancia con el estudio de Zagalaz (Zagalaz et al., 2021). Este trabajo mostró una sensibilidad y especificidad aceptable, aunque más baja que las anteriores (0.78- 0.79).

En la tabla 5 se muestran los resultados de sensibilidad, especificidad y valores predictivos para cada estudio.

Estudio	Año	Sensibilidad	IC 95%	Especificidad	IC 95%	VPP	VPN
Lövgren,	2016	0,81	0,73-0,87	0,79	0,73-0,83	0,3	0,97
Lövgren,	2018	0,96	0,92-0,98	0,34	0,28-0,40	0,59	0,9
Stasiak	2020	0,97	0,93-0,99	0,26	0,18-0,35	0,85	0,68
Kaynak	2020	0,94	0,86- 0,96	0,83	0,75-0,93	0,92	0,85
Zhang	2019	0,96	0,94- 0,97	0,72	0,59-0,83	0,97	0,62
Dos Santos	2014	0,86	0,78- 0,91	0,92	0,83-0,96	0,94	0,84
Yap	2021	0,95	0,93- 0,96	0,88	0,78-0,95	0,99	0,52
Zagalaz	2021	0,78	0,64-0,88	0,79	0,65-0,89	0,78	0,79
Yap	2021	0,95	0,93-0,96	0,93	0,86-0,97	0,99	0,42
Pires	2018	0,86	0,75-0,94	0,95	0,87-0,99	0,94	0,99

Tabla 5. Estudios 1y 2 (3Q-TMD). Estudios 3 al 7 (IAF). Estudios 8 al 10 (IAFC)

Síntesis de resultados

A pesar del escaso número de estudios 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 para todas las pruebas valores de sensibilidad muy buenos: 0,92 con intervalos de confianza

(IC) de 0,88 a 0,95. Por su parte los valores de especificidad mostraron una precisión cercana al 80%: 0.79 IC (0,63; 090).

Metaanálisis

Para ambas pruebas a evaluar se estimaron los puntos operativos resumidos (sensibilidades y especificidades resumidas), bajo un intervalo de confianza del 95 %. Para unificar en un solo resultado la PPD se calculó el ORD y se estimó un área bajo la curva utilizando el modelo de riesgo proporcionales.

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 importantes. Con respecto al test 3Q-TMD se pudieron evaluar solo 2 estudios uno de los cuales se realizó en un centro especializado de dolor facial lo que elevó los valores de sensibilidad, pero disminuyó significativamente los de especificidad. Con respecto a la prueba índice IAF casi todos los resultados presentaron una muy buena sensibilidad. Solo el estudio de Zagalaz mostró niveles de sensibilidad más pobres. El grafico 2 muestra los resultados del análisis de sensibilidad.

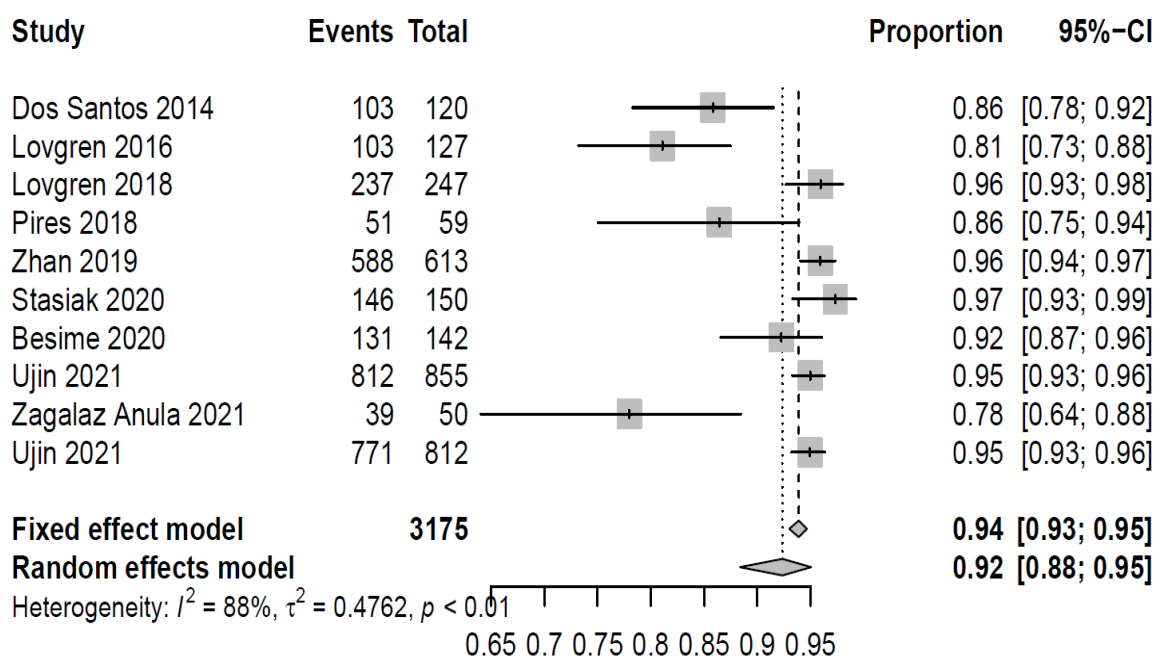
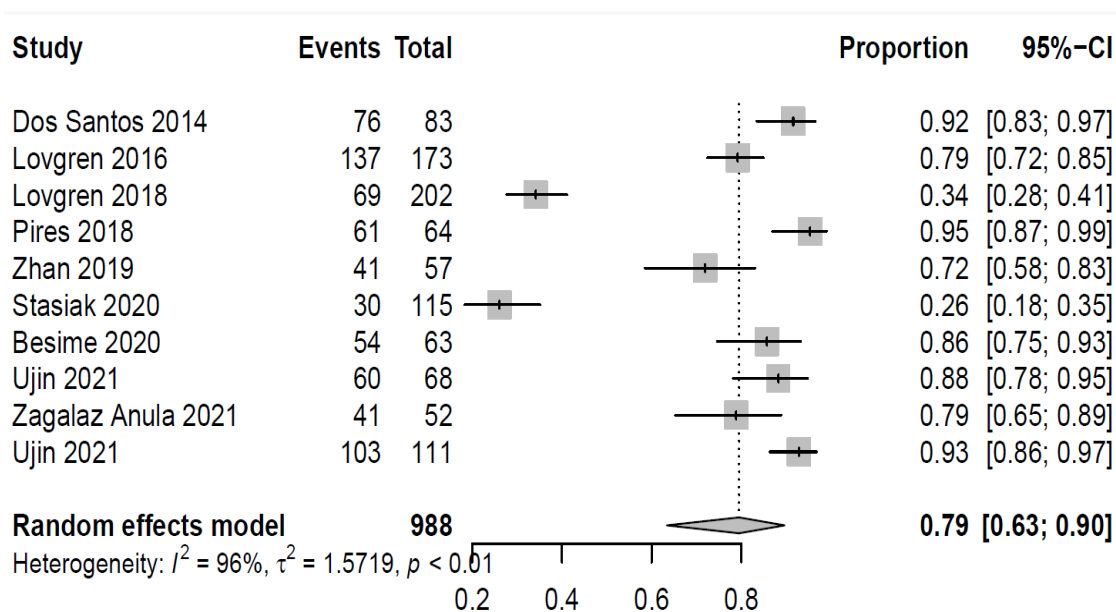


Gráfico 2. Diagrama de bosque de la sensibilidad

Especificidad

Una forma de definir la especificidad es la capacidad de una prueba para identificar los individuos sin la afección buscada. En el caso de 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). Los dos estudios que mostraron una menor especificidad de manera muy marcada fueron el de Lovgren (Lovgren et al., 2018) y el de Stasiak (Stasiak et al., 2020). El primero representa la prueba 3Q-TMD y el segundo la prueba IAF. (Gráfico 3)



Al observar la heterogeneidad tanto para la Sensibilidad como para la Especificidad, los estadísticos Q , τ y I sugieren la existencia de heterogeneidad entre estudios, con lo cual es recomendable usar un modelo de efectos aleatorios.

Gráfico 3. Diagrama de bosque de la especificidad

Medidas Adicionales

Razón de verosimilitud y Odds Ratio Diagnóstico

Tanto los valores de sensibilidad y especificidad como los valores predictivos estiman parcialmente la capacidad de una prueba diagnóstica, por lo que es necesario utilizarlos en parejas. Cuando se pretende buscar un solo valor que represente la precisión diagnóstica puede utilizarse el odds ratio diagnóstico (ORD), que no es más que la razón entre las chances de estar enfermo si la prueba da positivo y las chances de no estar

enfermo si la prueba da negativo. A las chances de estar enfermo cuando la prueba da positivo se lo conoce como likelihood ratio positivo (LR+) o razón de verosimilitud positiva, mientras que las chances de no estar enfermo si la prueba da negativo se conoce como likelihood ratio negativo (LR-) o razón de verosimilitud negativo. En síntesis, el ORD no es más que el cociente entre estas dos razones (LR+/LR-).

El ORD es una medida que permite sintetizar en un solo valor la performance diagnóstica de una prueba. Los valores de ORD varían de cero a infinito. El valor nulo es el uno, que significa que la prueba no tiene capacidad discriminatoria entre sanos y enfermos. Un valor mayor de uno indica capacidad discriminatoria, que será mayor cuanto mayor sea el valor. Por último, valores entre cero y uno indican que la prueba no discrimina bien entre enfermos y sanos. Observando la distribución espacial de los 10 estudios puede decirse que hay 2 de ellos que muestran una Tasa de falsos positivos grande, por arriba de 0.7, lo que habla de valores bajos de especificidad, en particular los estudios Lovgren 2018 y Stasiak 2020. El valor global de ORD para el conjunto de pruebas fue de (3.81) IC 95% (3.10; 4.53) observando la tabla con las estimaciones puntuales para el ORD, se puede decir que los tests bajo análisis se muestran como buenas herramientas de detección. (Gráfico 4)

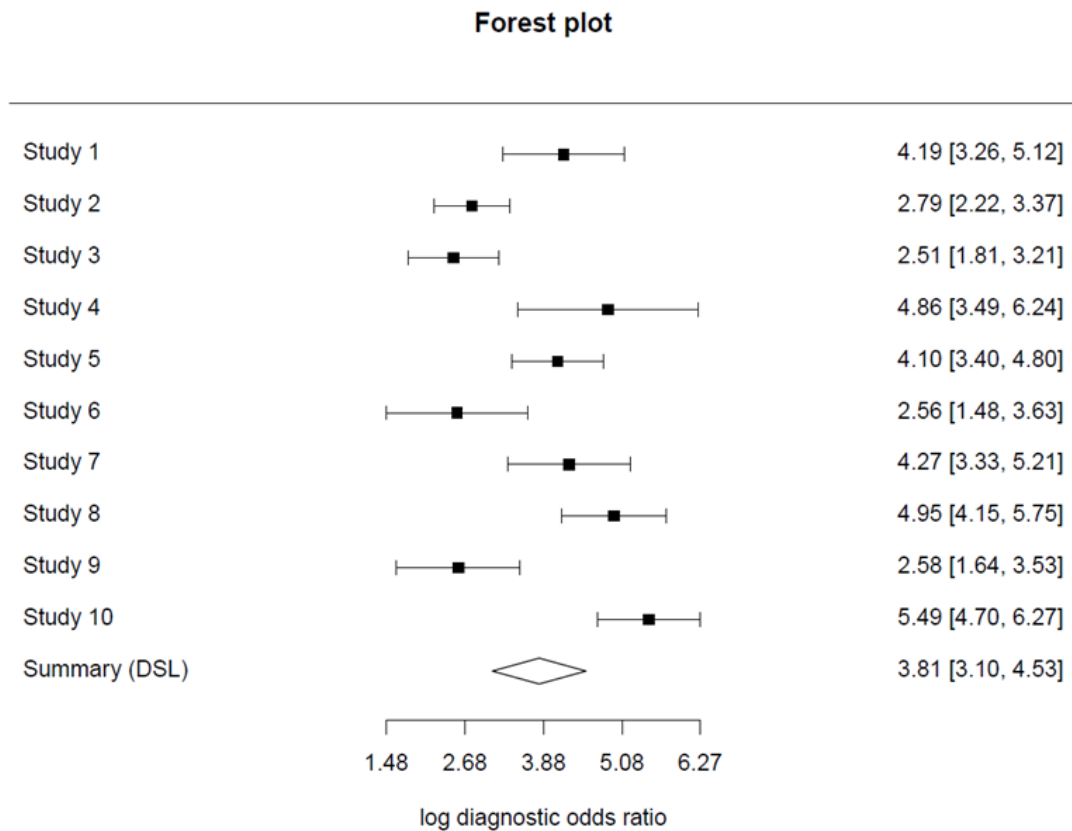


Gráfico 4. Diagrama de bosque del ODR (Odds Ratio Diagnostic)

La información que se muestra en el gráfico 4 es la base para el análisis estadístico de los estudios PPD en metanálisis donde es posible estimar una curva ROC resumida a través de los estudios observados.

Curvas de Rendimiento Diagnóstico

Las curvas de rendimiento diagnóstico o curvas ROC son una representación gráfica que relaciona la proporción de verdaderos positivos (sensibilidad) con la proporción de falsos positivos (1 menos especificidad) para los diferentes posibles valores de una prueba diagnóstica con el propósito de determinar qué valor discrimina mejor entre enfermos y no enfermos. La curva característica operativa del receptor se construye a partir de una gráfica de dispersión, cuyos ejes de ordenadas (y) y abscisas (x) corresponden respectivamente a la sensibilidad y al complemento de la especificidad para los diferentes posibles resultados de la prueba diagnóstica. Desde la esquina inferior izquierda y la esquina superior derecha del gráfico se traza una línea punteada que recibe el nombre de “diagonal de referencia” o “línea de no discriminación”. Esta diagonal de referencia corresponde a la representación teórica de una prueba diagnóstica que no discrimina enfermos de no enfermos (distribución idéntica de los resultados para ambos grupos) (Pérez et al., 2021). La curva característica operativa del receptor de la prueba diagnóstica del ejemplo presentado (gráfico 5) consta de dos ejes que varían de 0 a 1 (0 a 100%), la sensibilidad en el eje vertical (y) y el complemento de la especificidad en el eje horizontal (x). Cada punto rojo en la curva característica operativa del receptor son posibles puntos de corte. La línea diagonal en negro recibe el nombre de diagonal de referencia o línea de no discriminación.

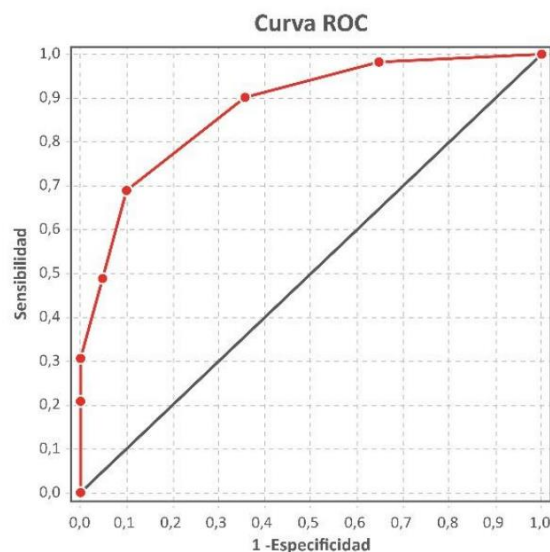


Gráfico 5. Curva de rendimiento diagnóstico

El área bajo la curva ROC es el indicador global de la precisión de una prueba diagnóstica. Esta área oscila entre 0,5 y 1. En el 1 las pruebas diagnósticas alcanzan el 100% de sensibilidad y de especificidad. Un área cercana a 0,5 significa que la prueba diagnóstica no discrimina pacientes enfermos de sanos.

Modelo de riesgos proporcionales

Para estimar el comportamiento de las pruebas sobre una curva ROC Se utilizó el enfoque del modelo de riesgos proporcionales. 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). Considerando el modelo bajo heterogeneidad los resultados fueron ligeramente menores, aunque manteniendo un desempeño muy bueno. AUC 0.94 IC 95% (0.97; 0.91).

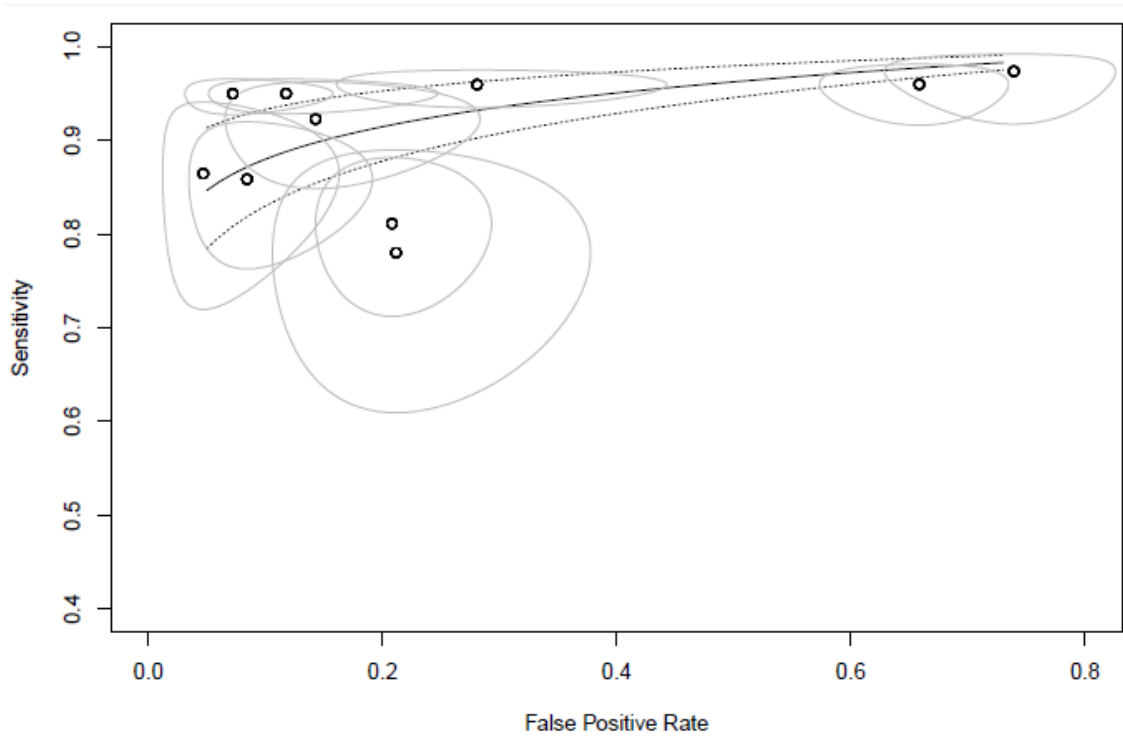


Gráfico 5: Curvas ROC

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.

Esta tesis 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 herramientas 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 (Dworkin & LeResche, 1992). 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 (Yap et al., 2021b). 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 en un 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 su puede deber al cuestionario de síntomas del DC-TMD (Lövgren, 2017).

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 (Giraudeau et al., 2017). 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 (Schiffman & Ohrbach, 2016). 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 (Pupo et al., 2016). 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 (Katzberg et al., 1996; Salé et al., 2013). 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. (Berni et al., 2015; Kaynak et al., 2020; Yap et al., 2021a). 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 de participantes, 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 (Rodrigues-Bigaton et al., 2017). 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 aun 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 especificidad. 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 (Manfredini et al., 2011). 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.

Podría hipotetizarse 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 como 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 (Talavera et al., 2011).

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 validados en su lengua original y en el caso del IAF solo en 3 idiomas más, debería alentarse a los investigadores a validar cualquiera de estos instrumentos en su propio idioma.

Limitaciones

En esta revisión sistemática podrían señalarse algunas limitaciones. En primer lugar, se detectó una gran heterogeneidad estadística entre los estudios. Si bien esto ocurre en la mayoría de los estudios sobre precisión de pruebas diagnósticas, se sumó el escaso número de artículos. Además, los estudios difieren principalmente en las características de la muestra y la heterogeneidad metodológica al comparar la variabilidad en el diseño de los estudios. Otro de los aspectos que se observó en esta revisión es el alto riesgo sesgo. En este sentido, de los 10 artículos publicados, 5 de ellos se consideraron con un alto riesgo de sesgo (Berni et al., 2015; Pires et al., 2018; Stasiak et al., 2020; Yap et al., 2021a; Zhang et al., 2020). Cuatro se consideraron de bajo riesgo (Kaynak et al., 2020; Lovgren et al., 2016, 2018; Yap et al., 2021b). Solamente uno se definió como no claro (Zhang et al., 2020).

Dentro de los aspectos metodológicos que influyeron en el riesgo de sesgo se pudo observar que la selección de pacientes en muchos de los estudios estaba basada en un diseño de “casos y controles”. Esta forma de selección de pacientes para la muestra ha

evidenciado que puede alterar la precisión diagnóstica de una prueba tal como lo advierten la herramienta Quadas2 y Prisma DTA.

Las limitaciones detectadas en el presente trabajo muestran la necesidad de implementar nuevas investigaciones con metodologías más estrictas y adecuadas a las preguntas de investigación.

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 odontológica.

Los resultados de la presente revisión sistemática y metaanálisis permiten concluir que tanto el IAF como EL 3Q-TMD son instrumentos altamente sensibles para la detección de TTM en pacientes.

Ambas pruebas de detección tienen la capacidad de identificar de manera rápida y simple 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 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 implementación 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

- Adler, R. H. (2009). Engel's biopsychosocial model is still relevant today. *Journal of Psychosomatic Research*, 67(6), 607–611.
<https://doi.org/10.1016/J.JPSYCHORES.2009.08.008>
- Aleem, I. S., Jalal, H., Aleem, I. S., Sheikh, A. A., & Bhandari, M. (2009). Clinical decision analysis: Incorporating the evidence with patient preferences. *Patient Preference and Adherence*, 3, 21. <https://doi.org/10.2147/PPA.S4549>
- Alex White, B., Williams, L. A., & Leben, J. R. (2001). Health Care Utilization and Cost among Health Maintenance Organization Members with Temporomandibular Disorders. *Journal of Orofacial Pain*, 15(2), 158–169.
- Ameer Al-Jundi, M., John, M. T., Setz, J. M., Szentpétery, A., & Kuss, O. (2008a). Meta-analysis of treatment need for temporomandibular disorders in adult nonpatients. *Journal of Orofacial Pain*, 22(2), 97–107.
<https://pubmed.ncbi.nlm.nih.gov/18548838/>
- Berni, K. C. dos S., Dibai-Filho, A. V., & Rodrigues-Bigaton, D. (2015b). Accuracy of the Fonseca anamnestic index in the identification of myogenous temporomandibular disorder in female community cases. *Journal of Bodywork and Movement Therapies*, 19(3), 404–409.
<https://doi.org/10.1016/J.JBMT.2014.08.001>
- Bevilaqua-Grossi, D., Chaves, T. C., De Oliveira, A. S., & Monteiro-Pedro, V. (2014). Anamnestic Index Severity and Signs and Symptoms of TMD. <https://doi.org/10.1179/CRN.2006.018>, 24(2), 112–118.
<https://doi.org/10.1179/CRN.2006.018>
- Bossuyt, P., & Leeflang, M. (2008). CHAPTER 6 DEVELOPING CRITERIA FOR INCLUDING STUDIES. In *Cochrane Handbook for Systematic Reviews of Diagnostic test Accuracy*.
- Breivik, H., Collett, B., Ventafridda, V., Cohen, R., & Gallacher, D. (2006). Survey of chronic pain in Europe: Prevalence, impact on daily life, and treatment. *European Journal of Pain*, 10(4), 287. <https://doi.org/10.1016/j.ejpain.2005.06.009>
- Ciapponi, A. (2015). *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*. 22–30. <http://www.bris.ac.uk/quadas>
- Cochrane, A. L., & Holland, W. W. (1971). *VALIDATION OF SCREENING PROCEDURES*.
- Dahlström, L., & Carlsson, G. E. (2010). Temporomandibular disorders and oral health-related quality of life. A systematic review. *Acta Odontologica Scandinavica*, 68(2), 80–85. <https://doi.org/10.3109/00016350903431118>
- Díaz, R. E. et al. (2009). Epidemiological study of chronic pain in Caldas, Colombia (DOLCA Study). *Acta Médica Colombiana*, 34, 96–102.

- Durham, J., Exley, C., Wassell, R., & Steele, J. G. (2007). "Management is a black art" - Professional ideologies with respect to temporomandibular disorders. *British Dental Journal*, 202(11). <https://doi.org/10.1038/bdj.2007.369>
- Durham, J., Steele, J. G., Wassell, R. W., & Exley, C. (2010). Living with uncertainty: Temporomandibular disorders. *Journal of Dental Research*, 89(8), 827–830. <https://doi.org/10.1177/0022034510368648>
- Durham, J., Steele, J., Moufti, M. A., Wassell, R., Robinson, P., & Exley, C. (2011). Temporomandibular disorder patients' journey through care. *Community Dentistry and Oral Epidemiology*, 39(6), 532–541. <https://doi.org/10.1111/j.1600-0528.2011.00608.x>
- Dworkin, S. F., & LeResche, L. (1992). Research diagnostic criteria for temporomandibular disorders: review, criteria, examinations and specifications, critique. *Journal of Craniomandibular Disorders : Facial & Oral Pain*, 6(4), 301–355.
- Epker, J., & Gatchel, R. J. (2000). *Coping Profile Differences in the Biopsychosocial Functioning of Patients With Temporomandibular Disorder*.
- Eusebi, P. (2013). E-Mail Methodological Notes Diagnostic Accuracy Measures. *Cerebrovasc Dis*, 36, 267–272. <https://doi.org/10.1159/000353863>
- fdi. (2020). *oral_health_definition-es*.
- Fernandes, G., van Selms, M. K. A., Gonçalves, D. A. G., Lobbezoo, F., & Camparis, C. M. (2015). Factors associated with temporomandibular disorders pain in adolescents. *Journal of Oral Rehabilitation*, 42(2), 113–119. <https://doi.org/10.1111/joor.12238>
- Ferrando, M., Andreu, Y., José Galdón, M., Durá, E., Poveda, R., & Vincente Bagán, J. (2004). Psychological variables and temporomandibular disorders: Distress, coping, and personality. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology and Endodontology*, 98(2), 153–160. <https://doi.org/10.1016/j.tripleo.2003.12.030>
- Fillington, R. B., Ohrbach, R., Greenspan, J. D., Knott, C., Diatchenko, L., Dubner, R., Bair, E., Baraian, C., MacK, N., Slade, G. D., & Maixner, W. (2013). Psychological Factors Associated With Development of TMD: The OPPERA Prospective Cohort Study. *The Journal of Pain*, 14(12), T75–T90. <https://doi.org/10.1016/J.JPAIN.2013.06.009>
- Finan, P. H., Goodin, B. R., & Smith, M. T. (2013). The association of sleep and pain: An update and a path forward. *The Journal of Pain : Official Journal of the American Pain Society*, 14(12), 1539. <https://doi.org/10.1016/J.JPAIN.2013.08.007>
- Giraudeau, A., Jeany, M., Ehrmann, E., Déjou, J., Ouni, I., & Orthlieb, J. D. (2017). Disc displacement without reduction: a retrospective study of a clinical diagnostic sign. *Cranio - Journal of Craniomandibular Practice*, 35(2), 86–93. <https://doi.org/10.1080/08869634.2016.1149291>

- Gonzalez, Y., Schiffman, E., & Gordon, S. (2011). Development of a brief and effective temporomandibular pain screening questionnaire: Reliability and Validity. *JADA*, 1183–1191.
- Griffiths, R. H. (1983). Report of the president's conference on the examination, diagnosis, and management of temporomandibular disorders. *Journal of the American Dental Association (1939)*, 106(1), 75–77.
<https://doi.org/10.14219/jada.archive.1983.0020>
- Health Information Research Unit - HIRU ~ Search Strategies for MEDLINE in Ovid Syntax and the PubMed translation.* (2022).
https://hiru.mcmaster.ca/hiru/HIRU_Hedges_MEDLINE_Strategies.aspx
- Helkimo, M. (1974). Studies on function and dysfunction of the masticatory system: IV. Age and sex distribution of symptoms of dysfunction of the masticatory system in lapps in the north of Finland. *Acta Odontologica Scandinavica*, 32(4), 255–267.
<https://doi.org/10.3109/00016357409026342>
- INFORM International Network for Orofacial Pain and Related Disorders Methodology.* (2023). buffalo.edu/rdc-tmdinternational/
- International Classification of Orofacial Pain, 1st edition (ICOP). (2020). *Cephalalgia : An International Journal of Headache*, 40(2), 129–221.
<https://doi.org/10.1177/0333102419893823>
- Katzberg, R. W., Westesson, P.-L., Tallents, R. H., Drake, C. M., & Davis, R. (1996). Orthodontics and temporomandibular joint internal derangement. In *Am J Orthod Dentofac Orthop* (Vol. 109).
- Kaynak, B. A., Taş, S., & Salkın, Y. (2020a). The accuracy and reliability of the Turkish version of the Fonseca anamnestic index in temporomandibular disorders. *Cranio : The Journal of Craniomandibular Practice*, 1–6.
<https://doi.org/10.1080/08869634.2020.1812808>
- Lipton, J. A., Ship, J. A., & Larach-Robinson, D. (1993). Estimated prevalence and distribution of reported orofacial pain in the United States. *Journal of the American Dental Association (1939)*, 124(10), 115–121.
<https://doi.org/10.14219/jada.archive.1993.0200>
- List, T. (2005). *Prevalence of Temporomandibular Pain and Subsequent Dental Treatment in Swedish Adolescents.* 19(2).
- List, T., & Jensen, R. H. (2017). Temporomandibular disorders: Old ideas and new concepts. In *Cephalalgia* (Vol. 37, Issue 7, pp. 692–704). SAGE Publications Ltd.
<https://doi.org/10.1177/0333102416686302>
- Lövgren, A. (2017). *Recognition of Temporomandibular Disorders.* UMEA.
- Lövgren, A., Parvaneh, H., Lobbezoo, F., Haggman-Henrikson, B., Wanman, A., & Visscher, C. M. (2018). Diagnostic accuracy of three screening questions (3Q/TMD) in relation to the DC/TMD in a specialized orofacial pain clinic. *ACTA*

ODONTOLOGICA SCANDINAVICA, 76(6), 380–386.
<https://doi.org/10.1080/00016357.2018.1439528>

- Lovgren, A., Visscher, C. M., Haggman-Henrikson, B., Lobbezoo, F., Marklund, S., & Wanman, A. (2016). Validity of three screening questions (3Q/TMD) in relation to the DC/TMD. *JOURNAL OF ORAL REHABILITATION*, 43(10), 729–736.
<https://doi.org/10.1111/joor.12428>
- Lundh, H., Westesson, P. L., & Kopp, S. (1987). A three-year follow-up of patients with reciprocal temporomandibular joint clicking. *Oral Surgery, Oral Medicine, Oral Pathology*, 63(5), 530–533. [https://doi.org/10.1016/0030-4220\(87\)90221-0](https://doi.org/10.1016/0030-4220(87)90221-0)
- Macfarlane, G. J. (2016). The epidemiology of chronic pain. *Pain*, 157(10), 2158–2159.
<https://doi.org/10.1097/J.PAIN.0000000000000676>
- Machado, L. P. e. S., Nery, C. de G., Leles, C. R., Nery, M. B. de M., & Okeson, J. P. (2009). The prevalence of clinical diagnostic groups in patients with temporomandibular disorders. *Cranio - Journal of Craniomandibular and Sleep Practice*, 27(3), 194–199. <https://doi.org/10.1179/crn.2009.029>
- Magnusson, T., Egermark, I., & Carlsson, G. E. (2000). A Longitudinal Epidemiologic Study of Signs and Symptoms of Temporomandibular Disorders from 15 to 35 Years of Age. *Journal of Orofacial Pain*, 14(4), 310–319.
- Magnusson, T., Egermark, I., & Carlsson, G. E. (2002). Treatment received, treatment demand, and treatment need for temporomandibular disorders in 35-year-old subjects. *Cranio : The Journal of Craniomandibular Practice*, 20(1), 11–17.
<https://doi.org/10.1080/08869634.2002.11746184>
- Manfredini, D., Guarda-Nardini, L., Winocur, E., Piccotti, F., Ahlberg, J., & Lobbezoo, F. (2011a). Research diagnostic criteria for temporomandibular disorders: A systematic review of axis i epidemiologic findings. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology and Endodontology*, 112(4), 453–462.
<https://doi.org/10.1016/j.tripleo.2011.04.021>
- Manfredini, D., Lombardo, L., & Siciliani, G. (2017). Dental occlusion and temporomandibular disorders. *Evidence-Based Dentistry*, 18(3), 86–87.
<https://doi.org/10.1111/joor.12531>
- Marklund, S., & Wänman, A. (2010). Risk factors associated with incidence and persistence of signs and symptoms of temporomandibular disorders. *Acta Odontologica Scandinavica*, 68(5), 289–299.
<https://doi.org/10.3109/00016357.2010.494621>
- McInnes, M. D. F., Moher, D., Thombs, B. D., McGrath, T. A., & Bossuyt, P. M. (2018). Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies The PRISMA-DTA Statement Supplemental content CME Quiz at jamanetwork.com/learning. *JAMA*, 319(4), 388–396.
<https://doi.org/10.1001/jama.2017.19163>
- McNeill, C., Danzig, W. M., Farrar, W. B., Gelb, H., Lerman, M. D., Moffett, B. C., Pertes, R., Solberg, W. K., Weinberg, L. A., & Walnut Creek, M. (1980).

MAXILLOFACIAL PROSTHETICS TEMPOROMANDIBULAR JOINT.
DENTAL IMPLANTS SECTION EDITORS Position paper of the American
Academy of Craniomandibular Disorders Craniomandibular (TMJ) disorders-The
state of the art. *Maxillofacial Prosthetics Temporomandibular Joint. Dental
Implants*, 434.

Melis, M., & Di Giosia, M. (2016). The role of genetic factors in the etiology of
temporomandibular disorders: A review. In *Cranio - Journal of Craniomandibular
Practice* (Vol. 34, Issue 1, pp. 43–51). Maney Publishing.
<https://doi.org/10.1179/2151090314Y.0000000027>

Nasstrom, A., Fallgren, J., Wanman, A., & Lovgren, A. (2019). The implementation of
a decision-tree did not increase decision-making in patients with
temporomandibular disorders in the public dental health service. *ACTA
ODONTOLOGICA SCANDINAVICA*, 77(5), 394–399.
<https://doi.org/10.1080/00016357.2019.1577989>

National Institute of Dental and Craniofacial Research. *Facial Pain*.
<http://www.nidcr.nih.gov/DataStatistics/FindDataByTopic/FacialPain/>. (n.d.).
Retrieved May 17, 2022, from [https://www.nidcr.nih.gov/research/data-
statistics/facial-pain](https://www.nidcr.nih.gov/research/data-statistics/facial-pain)

Nilsson, I., List, T., & Drangsholt, M. (2006). The Reliability and Validity of Self-
reported Temporomandibular Disorder Pain in Adolescents. *Journal of Orofacial
Pain*.
[https://web.p.ebscohost.com/abstract?direct=true&profile=ehost&scope=site&auth
type=crawler&jrnl=10646655&AN=36831130&h=5L5CIuO%2fWSNTZ4NmeCD
Poye%2fB18xrx4ApzL4WI9HOVz5bBe7HjS%2fo2H%2byf%2fgNJJMTrYGzdG
FzSJUAQvZAJaSiw%3d%3d&crl=c&resultNs=AdminWebAuth&resultLocal=Err
CrlNotAuth&crlhashurl=login.aspx%3fdirect%3dtrue%26profile%3dehost%26sco
pe%3dsite%26authtype%3dcrawler%26jrnl%3d10646655%26AN%3d36831130](https://web.p.ebscohost.com/abstract?direct=true&profile=ehost&scope=site&auth
type=crawler&jrnl=10646655&AN=36831130&h=5L5CIuO%2fWSNTZ4NmeCD
Poye%2fB18xrx4ApzL4WI9HOVz5bBe7HjS%2fo2H%2byf%2fgNJJMTrYGzdG
FzSJUAQvZAJaSiw%3d%3d&crl=c&resultNs=AdminWebAuth&resultLocal=Err
CrlNotAuth&crlhashurl=login.aspx%3fdirect%3dtrue%26profile%3dehost%26sco
pe%3dsite%26authtype%3dcrawler%26jrnl%3d10646655%26AN%3d36831130)

Oakley, C., & Brunette, D. M. (2002). *EVIDENCE BASED DENTISTRY THE USE OF
DIAGNOSTIC DATA IN CLINICAL DENTAL PRACTICE*.

Okeson, J. P. (2019). *Management of Temporomandibular Disorders and Occlusion*
(8^o). Elsevier Inc.

Pastore, G. P., Goulart, D. R., Pastore, P. R., Prati, A. J., & de Moraes, M. (2018).
Comparison of instruments used to select and classify patients with
temporomandibular disorder TT - Avaliação de instrumentos utilizados para
selecionar pacientes com disfunção temporomandibular. *Acta Odontológica
Latinoamericana*, 31(1), 16–22.
[http://www.scielo.org.ar/scielo.php?script=sci_arttext&pid=S1852-
48342018000100003&lang=pt](http://www.scielo.org.ar/scielo.php?script=sci_arttext&pid=S1852-
48342018000100003&lang=pt)

Pérez, I., Taito-Vicenti, I. Y., González-Xuriguera, C. G., Carvajal, C., Franco, J. V. A.,
& Loézar, C. (2021). How to interpret diagnostic tests. *Medwave*, 21(7).
<https://doi.org/10.5867/medwave.2021.07.8432>

- Pires, P. F., de Castro, E. M., Pelai, E. B., de Arruda, A. B. C., & Rodrigues-Bigaton, D. (2018). Analysis of the accuracy and reliability of the Short-Form Fonseca Anamnestic Index in the diagnosis of myogenous temporomandibular disorder in women. *Brazilian Journal of Physical Therapy*, 22(4), 276–282. <https://doi.org/10.1016/j.bjpt.2018.02.003>
- Progiante, P., Pattussi, M., Lawrence, H., Goya, S., Grossi, P., & Grossi, M. (2015). Prevalence of Temporomandibular Disorders in an Adult Brazilian Community Population Using the Research Diagnostic Criteria (Axes I and II) for Temporomandibular Disorders (The Maringá Study). *The International Journal of Prosthodontics*, 28(6), 600–609. <https://doi.org/10.11607/ijp.4026>
- Pupo, Y. M., Pantoja, L. L. Q., Veiga, F. F., Stechman-Neto, J., Zwir, L. F., Farago, P. V., De Luca Canto, G., & Porporatti, A. L. (2016). Diagnostic validity of clinical protocols to assess temporomandibular disk displacement disorders: a meta-analysis. In *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology* (Vol. 122, Issue 5, pp. 572–586). Mosby Inc. <https://doi.org/10.1016/j.oooo.2016.07.004>
- RAE. (2023). dle.rae.es/cribado
- Raja, S. N., Carr, D. B., Cohen, M., Finnerup, N. B., Flor, H., Gibson, S., Keefe, F., Mogil, J. S., Ringkamp, M., Sluka, K. A., Song, X.-J., Stevens, B., Sullivan, M., Tutelman, P., Ushida, T., & Vader, K. (2020). The Revised IASP definition of pain: concepts, challenges, and compromises HHS Public Access. *Pain*, 161(9), 1976–1982. <https://doi.org/10.1097/j.pain.0000000000001939>
- Rodrigues-Bigaton, D., de Castro, E. M., & Pires, P. F. (2017). Factor and Rasch analysis of the Fonseca anamnestic index for the diagnosis of myogenous temporomandibular disorder. *Brazilian Journal of Physical Therapy*, 21(2), 120–126. <https://doi.org/10.1016/J.BJPT.2017.03.007>
- Salameh, J. P., Bossuyt, P. M., McGrath, T. A., Thombs, B. D., Hyde, C. J., MacAskill, P., Deeks, J. J., Leeftang, M., Korevaar, D. A., Whiting, P., Takwoingi, Y., Reitsma, J. B., Cohen, J. F., Frank, R. A., Hunt, H. A., Hooft, L., Rutjes, A. W. S., Willis, B. H., Gatsonis, C., ... McInnes, M. D. F. (2020). Preferred reporting items for systematic review and meta-analysis of diagnostic test accuracy studies (PRISMA-DTA): Explanation, elaboration, and checklist. In *The BMJ* (Vol. 370, Issue 12). BMJ Publishing Group. <https://doi.org/10.1136/bmj.m2632>
- Salé, H., Bryndahl, F., & Isberg, A. (2013). Temporomandibular joints in asymptomatic and symptomatic nonpatient volunteers: A prospective 15-year follow-up clinical and MR imaging study. *Radiology*, 267(1), 183–194. <https://doi.org/10.1148/radiol.12112243>
- Sanders, A. E., Slade, G. D., Bair, E., Fillingim, R. B., Knott, C., Dubner, R., Greenspan, J. D., Maixner, W., & Ohrbach, R. (2013). General health status and incidence of first-onset temporomandibular disorder: The OPPERA prospective cohort study. *Journal of Pain*, 14(12 SUPPL.). <https://doi.org/10.1016/j.jpain.2013.06.001>

- Schiffman, E. L., Professor, Msa., Truelove, E. L., Ohrbach, R., Professor, P., Anderson, G. C., John, M. T., List, T., Look, J. O., & Research Associate, Mphs. (2010). Assessment of the Validity of the Research Diagnostic Criteria for Temporomandibular Disorders: Overview and Methodology. *J Orofac Pain*, 24(1), 7–24.
- Schiffman, E., & Ohrbach, R. (2016). Executive summary of the Diagnostic Criteria for Temporomandibular Disorders for clinical and research applications. *Journal of the American Dental Association*, 147(6), 438–445. <https://doi.org/10.1016/j.adaj.2016.01.007>
- Schiffman, E., Ohrbach, R., Truelove, E., Look, J., Anderson, G., Goulet, J.-P., List, T., Svensson, P., Gonzalez, Y., Lobbezoo, F., Michelotti, A., Brooks, S. L., Ceusters, W., Drangsholt, M., Ettl, D., Gaul, C., Goldberg, L. J., Haythornthwaite, J. A., Hollender, L., ... Dworkin, S. F. (2014). Diagnostic Criteria for Temporomandibular Disorders (DC/TMD) for Clinical and Research Applications: Recommendations of the International RDC/TMD Consortium Network* and Orofacial Pain Special Interest Group†. *Journal of Oral & Facial Pain and Headache*, 28(1), 6–27. <https://doi.org/10.11607/jop.1151>
- Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L. A., Altman, D. G., Booth, A., Chan, A. W., Chang, S., Clifford, T., Dickersin, K., Egger, M., Gøtzsche, P. C., Grimshaw, J. M., Groves, T., Helfand, M., ... Whitlock, E. (2015). Preferred reporting items for systematic review and meta-analysis protocols (prisma-p) 2015: Elaboration and explanation. In *BMJ (Online)* (Vol. 349). BMJ Publishing Group. <https://doi.org/10.1136/bmj.g7647>
- Stasiak, G., Maracci, L. M., de Oliveira Chami, V., Pereira, D. D., Tomazoni, F., Bernardon Silva, T., Ferrazzo, V., & Marquezan, M. (2020). TMD diagnosis: Sensitivity and specificity of the Fonseca Anamnestic Index. *Cranio : The Journal of Craniomandibular Practice*, 1–5. <https://doi.org/10.1080/08869634.2020.1839724>
- Stowell, A. W., Wildenstein, L., & Gatchel, R. J. (2007). Cost-effectiveness of treatments for temporomandibular disorders: Biopsychosocial intervention versus treatment as usual. *Journal of the American Dental Association*, 138(2), 202–208. <https://doi.org/10.14219/jada.archive.2007.0137>
- Suvinen, T. I., Reade, P. C., Kemppainen, P., Könönen, M., & Dworkin, S. F. (2005). Review of aetiological concepts of temporomandibular pain disorders: Towards a biopsychosocial model for integration of physical disorder factors with psychological and psychosocial illness impact factors. *European Journal of Pain*, 9(6), 613. <https://doi.org/10.1016/j.ejpain.2005.01.012>
- Talavera, J. O., Wachter-Rodarte, N. H., Rivas-Ruiz, R., Nacional, C. M., & Xxi, S. (2011). Estudios de proceso (prueba diagnóstica). In *Rev Med Inst Mex Seguro Soc* (Vol. 49, Issue 2).
- Türp, J. C., & Schindler, H. (2012). The dental occlusion as a suspected cause for TMDs: Epidemiological and etiological considerations. In *Journal of Oral*

Rehabilitation (Vol. 39, Issue 7, pp. 502–512). <https://doi.org/10.1111/j.1365-2842.2012.02304.x>

- Valesan, L. F., Da-Cas, C. D., Réus, J. C., Denardin, A. C. S., Garanhani, R. R., Bonotto, D., Januzzi, E., & de Souza, B. D. M. (2021). Prevalence of temporomandibular joint disorders: a systematic review and meta-analysis. In *Clinical Oral Investigations* (Vol. 25, Issue 2, pp. 441–453). Springer Science and Business Media Deutschland GmbH. <https://doi.org/10.1007/s00784-020-03710-w>
- Van den Berghe, L. I., De Clercq, E., & Marks, L. A. (2017). Treatment needs and therapy experiences in patients with temporomandibular disorders: a retrospective survey. *Cranio - Journal of Craniomandibular Practice*, 35(2), 116–121. <https://doi.org/10.1080/08869634.2016.1159814>
- Wanman, A., & Agerberg, G. (1986). *Two-year longitudinal study of symptoms of mandibular dysfunction in adolescents*.
- Wänman, A., & Wigren, L. (1995). 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 Odontologica Scandinavica*, 53(5), 318–324. <https://doi.org/10.3109/00016359509005994>
- Whiting, P. F., Rutjes, A. W. S., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., Leeflang, M. M. G., Sterne, J. A. C., & Bossuyt, P. M. M. (2011). Quadas-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine*, 155(8), 529–536. <https://doi.org/10.7326/0003-4819-155-8-201110180-00009>
- Willeman, L. V., Tesch, B., De, R., Tesch, S., Pereira, F. J., & Bastos, W. (2014). Trastornos temporomandibulares y dolor orofacial crónico: al final, ¿a qué área pertenecen? In *Rev Soc Esp Dolor* (Vol. 21, Issue 2).
- Williams, A. C. D. C., & Craig, K. D. (2016). Updating the definition of pain. In *Pain* (Vol. 157, Issue 11, pp. 2420–2423). Lippincott Williams and Wilkins. <https://doi.org/10.1097/j.pain.0000000000000613>
- Wirz, S., Ellerkmann, R. K., Buecheler, M., Putensen, C., Nadstawek, J., Wartenberg, H.-C., & Hospitals, C. (2010). *Management of Chronic Orofacial Pain: A Survey of General Dentists in German University Hospitals*. 416–424. <https://academic.oup.com/painmedicine/article/11/3/416/1886127>
- Yap, A. U., Zhang, M. J., Lei, J., & Fu, K. Y. (2021a). Accuracy of the Fonseca Anamnestic Index for identifying pain-related and/or intra-articular Temporomandibular Disorders. *Cranio - Journal of Craniomandibular Practice*. <https://doi.org/10.1080/08869634.2021.1954375>
- Yap, A. U., Zhang, M. J., Lei, J., & Fu, K. Y. (2021b). Diagnostic accuracy of the short-form Fonseca Anamnestic Index in relation to the Diagnostic Criteria for Temporomandibular Disorders. *Journal of Prosthetic Dentistry*. <https://doi.org/10.1016/j.prosdent.2021.02.016>

Zagalaz-Anula, N., Sanchez-Torrelo, C., & Acebal-Blanco, F. (2021). The Short Form of the Fonseca Anamnestic Index for the Screening of Temporomandibular Disorders_ Validity and Reliability in a Spanish-Speaking Population _ Enhanced Reader. *Journal of Clinical Medicine*.

Zhang, M. Juan, Yap, A. U. J., Lei, J., & Fu, K. Y. (2020a). Psychometric evaluation of the Chinese version of the Fonseca anamnestic index for temporomandibular disorders. *Journal of Oral Rehabilitation*, 47(3), 313–318.
<https://doi.org/10.1111/joor.12893>