



UCAM

UNIVERSIDAD CATÓLICA
DE MURCIA

ESCUELA INTERNACIONAL DE DOCTORADO
Programa de Doctorado en Ciencias de la Salud

Validación de herramientas de evaluación
en terapia ocupacional y enfermería a través
de la metodología COSMIN.

Autor:

D. Pedro Montagut Martínez

Directores:

Dr. D. David Pérez Cruzado

Dr. D. José Joaquín García Arenas

Murcia, 27 de mayo de 2022



UCAM

UNIVERSIDAD CATÓLICA
DE MURCIA

ESCUELA INTERNACIONAL DE DOCTORADO Programa de Doctorado en Ciencias de la Salud

Validación de herramientas de evaluación
en terapia ocupacional y enfermería a través
de la metodología COSMIN.

Autor:

D. Pedro Montagut Martínez

Directores:

Dr. D. David Pérez Cruzado

Dr. D. José Joaquín García Arenas

Murcia, 27 de mayo de 2022



UCAM

UNIVERSIDAD CATÓLICA
DE MURCIA

**AUTORIZACIÓN DEL DIRECTOR DE LA TESIS
PARA SU PRESENTACIÓN**

El Dr. D. David Pérez Cruzado y el Dr. D. José Joaquín García Arenas como Directores⁽¹⁾ de la Tesis Doctoral titulada “Validación de herramientas de evaluación en terapia ocupacional y enfermería a través de la metodología COSMIN” realizada por D. Pedro Montagut Martínez en el Programa de Doctorado en Ciencias de la Salud, **autorizan su presentación a trámite** dado que reúne las condiciones necesarias para su defensa.

Lo que firmo, para dar cumplimiento al Real Decreto 99/2011 de 28 de enero, en Murcia a 27 de Mayo de 2022.

Fdo: Dr. D. David Pérez Cruzado

Fdo: Dr. D. José Joaquín García Arenas

⁽¹⁾ Si la Tesis está dirigida por más de un Director tienen que constar y firmar ambos.

UCAM



EIDUCAM
Escuela Internacional
de Doctorado

AGRADECIMIENTOS

En primer lugar, me gustaría agradecer a mis directores la confianza depositada en mí para la realización de este proyecto.

A nivel individual, quería mostrar mi gratitud a José Joaquín García por facilitarme la entrada en este programa de Doctorado en Ciencias de la Salud y por la gran experiencia vivida durante las prácticas en el centro de salud mental de Lorca, cuando era alumno de terapia ocupacional. Gracias por haber hecho que mi vida fuera más fácil durante ese tiempo. Agradecerte también la dedicación y todos los consejos dados durante mi formación como doctorando.

A David Pérez, gracias por todo tu tiempo (que ha sido mucho, infinito), por tu paciencia, compromiso, apoyo constante y disponibilidad. Por compartir muchos de tus conocimientos sobre investigación y por enseñarme a querer ser parte de este mundo de la investigación.

Ambos sois un ejemplo a nivel profesional y personal y un reflejo de lo que algún día pretendo ser.

En segundo lugar, quería dar las gracias a mi familia. A mi madre que siempre dice que esta muy orgullosa de sus hijos; a mi padre, que aunque ya no esta para compartir mis alegrías, se que estaría muy feliz de lo logrado por todos nosotros; a mi mujer, por su paciencia, apoyo y comprensión en los momentos difíciles y a mi hija, que espero que algún día cuando mire al pasado pensando en su padre pueda ver mi reflejo en su futuro.

A mis padres. A mis hermanas.

A mi mujer Carmen. A mi hija Jimena.

“Hay una fuerza más poderosa que el vapor, la electricidad y la energía atómica: la voluntad”

-Albert Einstein-

ÍNDICE GENERAL

I. INTRODUCCIÓN.....	35
II. JUSTIFICACIÓN.....	41
III. MARCO TEÓRICO.....	47
1. CUESTIONARIOS DE EVALUACIÓN EN CIENCIAS DE LA SALUD.....	51
1.1 DEFINICIÓN DE CUESTIONARIO.....	51
1.2 PRINCIPIOS Y PROPIEDADES DE UN INSTRUMENTO DE MEDIDA.....	52
1.3 PROCESO PARA EL DESARROLLO DE UN CUESTIONARIO.....	52
1.3.1 Definición del constructo o aspecto que pretendemos medir.....	52
1.3.2 Finalidad del cuestionario.....	53
1.3.3 Elaboración de los ítems.....	54
1.3.4 Número de ítems.....	54
1.3.5 Contenido.....	54
1.3.6 Definición y disposición.....	54
1.3.7 Identificación de los sesgos al completar el cuestionario.....	55
1.3.8 Clasificación de las respuestas.....	55
1.3.9 Puntuación de los ítems.....	55
1.3.10 Fase de validación.....	56
1.3.11 Prueba piloto o pre-test.....	56
1.3.12 Evaluación de las propiedades psicométricas del cuestionario.....	56

1.4	CAUSAS QUE PUEDEN ALTERAR TANTO LA VALIDEZ COMO LA FIABILIDAD DE UN CUESTIONARIO.....	59
1.5	VALIDEZ Y ADAPTACIÓN DE CUESTIONARIOS.....	60
2.	PROPIEDADES PSICOMÉTRICAS DE LOS CUESTIONARIOS.....	63
2.1	VALIDEZ DE CONTENIDO.....	64
2.2	VALIDEZ ESTRUCTURAL.....	68
2.3	CONSISTENCIA INTERNA.....	70
2.4	FIABILIDAD.....	71
2.4.1	Fiabilidad intra-observador o fiabilidad test-retest.....	72
2.4.2	Fiabilidad inter-observador.....	72
2.5	ERROR DE MEDICIÓN.....	73
2.5.1	Parámetros estimados del SEM.....	73
2.5.2	Etapas del SEM.....	74
2.6	VALIDEZ DE CRITERIO CONVERGENTE.....	75
2.7	VALIDEZ DE CRITERIO.....	76
2.8	PRUEBA DE HIPÓTESIS.....	77
2.9	SENSIBILIDAD AL CAMBIO.....	77
2.10	VALIDEZ TRANSCULTURAL.....	78
3.	NORMATIVA COSMIN.....	83
3.1	LISTA DE VERIFICACIÓN COSMIN.....	83
3.2	GUÍA COSMIN PARA REVISIONES SISTEMÁTICAS.....	85

(A. Búsqueda de la literatura.	85
3.2.1 Punto 1. Formular el objetivo de la revisión.....	85
3.2.2 Punto 2. Establecer los criterios de elegibilidad.....	85
3.2.3 Punto 3. Realizar una búsqueda bibliográfica.....	86
3.2.4 Punto 4. Seleccionar resúmenes y artículos a texto completo.....	88
(B. Evaluación de las propiedades de medición.....	89
3.2.5 Punto 5. Evaluación de la validez de contenido.	90
3.2.6 Punto 6. Evaluación de la estructura interna: validez estructural, consistencia interna y validez transcultural / invariancia de medición.	91
3.2.7 Punto 7. Evaluación del resto de propiedades.....	92
(C. Selección del cuestionario	92
3.2.8 Punto 8. Descripción de la interpretabilidad y la viabilidad.	92
3.2.9 Punto 9. Formulación de recomendaciones.....	93
3.2.10 Punto 10. Información sobre la revisión sistemática.....	93
4. CRITERIOS DE TERWEE Y PRINSEN.....	97
4.1 PROPIEDADES PSICOMÉTRICAS: CALIDAD DE LOS RESULTADOS.....	100
4.1.1 Validez de contenido	100
4.1.2 Validez estructural	106
4.1.3 Consistencia interna.....	107
4.1.4 Fiabilidad.....	107
4.1.5 Error de medición.....	108
4.1.6 Prueba de hipótesis para la validez de constructo	108

16

4.1.7	Validez transcultural / invariancia de medición.....	109
4.1.8	Validez de criterio	109
4.1.9	Sensibilidad al cambio	110
5.	MÉTODO GRADE.....	113
5.1	EVALUACIÓN DE LA CALIDAD DE LA EVIDENCIA Y ELABORACIÓN DE LA RECOMENDACIÓN.....	114
6.	LISTA DE VERIFICACIÓN STARD.....	121
6.1	ÍTEMS.....	122
6.1.1	Título / resumen / palabras clave	122
6.1.2	Introducción.....	122
6.1.3	Métodos	122
6.1.4	Resultados	124
6.1.5	Discusión	125
7.	LISTA DE VERIFICACIÓN STROBE.....	129
7.1	ÍTEMS.....	130
7.1.1	Título y Resumen. Ítem 1	130
7.1.2	Introducción. Ítems 2 y 3.....	130
7.1.3	Metodología. Ítems 4-12	130
7.1.4	Resultados. Ítems 13-17	131
7.1.5	Discusión. Ítems 18-21	131
7.1.6	Financiación y patrocinio. Ítem 22.....	132
8.	POBLACIONES DE ESTUDIO	137

8.1	DIABETES	137
8.1.1	Definición, tipos de diabetes y diagnóstico	137
8.1.2	Epidemiología y carga global de la diabetes	138
8.1.3	Etiopatología y consecuencias de la diabetes.....	138
8.1.4	Tratamiento no farmacológico	139
8.1.5	Tratamiento farmacológico	141
8.1.6	Autocontrol	143
8.1.7	Cambios en el estilo de vida	143
8.2	CÁNCER	144
8.2.1	Definición	144
8.2.2	Epidemiología.....	144
8.2.3	Etiología y factores de riesgo	144
8.2.4	Prevención frente al cáncer	145
8.2.5	Detección temprana	145
8.2.6	Tratamiento.....	147
8.2.7	Cuidados paliativos	147
8.2.8	Eutanasia	148
8.3	DISFUNCIÓN DEL PROCESAMIENTO SENSORIAL	148
8.3.1	Definición	148
8.3.2	Etiología y consecuencias.....	149
8.3.3	Disfunciones de la integración sensorial.....	149

8.3.4	Evaluación en integración sensorial	151
8.3.5	Intervención en integración sensorial.....	153
IV. OBJETIVOS	155
V. MARCO EMPÍRICO	159
ARTÍCULO 1. NUTRITIONAL STATUS MEASUREMENT INSTRUMENTS: A SYSTEMATIC PSYCHOMETRIC REVIEW	163
ARTÍCULO 2. THE DIETARY KNOWLEDGE MEASUREMENT INSTRUMENTS IN DIABETES: A SYSTEMATIC PSYCHOMETRIC REVIEW	203
ARTÍCULO 3. CANCER-RELATED FATIGUE MEASURES IN PALLIATIVE CARE: A PSYCHOMETRIC SYSTEMATIC REVIEW	267
ARTÍCULO 4. VALIDATION OF THE ADOLESCENT/ADULT SENSORY PROFILE (AASP) FOR SPAIN	387
VI. DISCUSIÓN	419
METODOLOGÍA COSMIN PARA REVISIONES SISTEMÁTICAS.....		421
METODOLOGÍA COSMIN PARA VALIDACIÓN DE ESCALAS.....		434
LIMITACIONES.....		442
IMPLICACIONES PARA LA PRÁCTICA CLÍNICA, PARA LA INVESTIGACIÓN Y RECOMENDACIONES PARA FUTURAS LÍNEAS DE INVESTIGACIÓN		445
VII. CONCLUSIONES	449
VIII. BIBLIOGRAFÍA	453
ANEXOS	503

ÍNDICE DE TABLAS

Tabla 1. Estrategia de búsqueda.....	167
Tabla 2. Características y descripción de los instrumentos de medición del estado nutricional en DM.....	173
Tabla 3. Calidad metodológica de los estudios y calidad de los resultados informados por propiedad de medición, instrumento y estudio.....	184
Tabla 4. Resumen de hallazgos por propiedad de medición e instrumento.....	186
Tabla 5. Estrategia de búsqueda.....	209
Tabla 6. Características de los estudios y descripción de los instrumentos de conocimiento de la diabetes.....	219
Tabla 7. Marco teórico y proceso de desarrollo del instrumento.....	243
Tabla 8. Calidad metodológica de los estudios y calidad de los resultados informados por propiedad de medición, instrumento y estudio.....	249
Tabla 9. Resumen de hallazgos: calidad metodológica, calidad de los resultados y calidad de la evidencia reportada por propiedad e instrumento de medición. ...	252
Tabla 10. Estrategia de búsqueda.....	275
Tabla 11. Calidad de los estudios incluidos (Strobe Checklist).....	284

Tabla 12. Instrumentos identificados para evaluar la fatiga en personas con cáncer en cuidados paliativos.	288
Tabla 13. Características de los estudios y descripción de las medidas de fatiga relacionada con el cáncer en los instrumentos de cuidados paliativos.	290
Tabla 14. Calidad metodológica de los estudios y calidad de los resultados informados por propiedad de medición, instrumento y estudio.	344
Tabla 15. Resumen de hallazgos: calidad metodológica, calidad de los resultados y calidad de la evidencia reportada por propiedad e instrumento de medición.	350
Tabla 16. Características de los participantes.	394
Tabla 17. Matriz de carga de factores después de la rotación varimax.	396
Tabla 18. Estadística descriptiva y consistencia interna para los ítems de la AASP.	400
Tabla 19. Fiabilidad test-retest y error de medición para el total de la muestra.	403
Tabla 20. Validez de criterio convergente de AASP- 42 ítems.	405

ÍNDICE DE FIGURAS

Figura 1. Scree plot (gráfico de pantalla de factores rotados).....	69
Figura 2. Análisis factorial confirmatorio (CFA) a través del programa de ecuaciones estructurales SPSS AMOS.	70
Figura 3. Estrategia de búsqueda en Pubmed.....	87
Figura 4. Diagrama de flujo COSMIN del proceso de búsqueda.....	88
Figura 5. Método GRADE modificado para estudios de validación.	90
Figura 6. Criterios de Terwee y Prinsen.	99
Figura 7. Guía de los 10 criterios para una buena validez de contenido.	103
Figura 8. Criterios COSMIN y sistema de calificación para evaluar la validez de contenido.....	104
Figura 9. Clasificación de la calidad de la evidencia sobre la validez de contenido (método GRADE modificado).....	105
Figura 10. Definición de los niveles de la calidad de la evidencia en el método GRADE.....	114
Figura 11. Factores que disminuyen y aumentan la calidad de la evidencia en GRADE.....	117

Figura 12. Lista de verificación STARD.....	126
Figura 13. Lista de verificación STROBE.....	134
Figura 14. Algoritmo para el tratamiento de la diabetes mellitus 2.....	142
Figura 15. Diagrama de flujo de la selección del estudio artículo 1.....	171
Figura 16. Diagrama de flujo de la selección del estudio artículo 2.....	214
Figura 17. Diagrama de flujo de la selección del estudio artículo 3.....	287
Figura 18. Scree plot (gráfico de pantalla de factores rotados).....	398
Figura 19. Gráfica análisis factorial confirmatorio (CFA).	402

ABREVIATURAS

(3LNQ) Three Levels of Needs Questionnaire

(AASP) Adolescent/Adult Sensory Profile

(AATM) Agencia de Evaluación de Tecnología Médica

(ACCF/AHA) American College of Cardiology Foundation / American Heart Association

(ADA) American Diabetes Association

(ADF) Distribución asintótica libre

(ADULT-SI) Adult Sensory Interview

(AHRQ) Agency for Healthcare Research and Quality

(ANOVA) Análisis de la varianza

(APA) American Psychological Association

(AQEL) Palliative Care Quality of Life Assessment

(ASI) Ayres Sensory Integration

(ASP) Adult Sensory Profile

(ASQ) Adult Sensory Questionnaire

(AUC) Área bajo la curva

(AVD) Actividades de la vida diaria

(CACG) Canadian Asthma Consensus Group

(CAMPAS-R) Cambridge Palliative Assessment Schedule

(CEBM) Centre for Evidence-Based Medicine

(CFI) Índice de ajuste comparativo

(CFS) Cancer Fatigue Scale

(CIIC) Centro Internacional de Investigaciones sobre el Cáncer

(CONSORT) CONSolidated Standards of Reporting Trials

(COS) Conjunto mínimo de resultados acordados

(COSMIN) COnsensus-based Standards for the selection of health status Measurement Instruments / Estándares basados en el COnsenso para la selección de INstrumentos de Medición de la salud

(CP) Cuidados paliativos

(C-SDSCA) Chinese- Summary of Diabetes Self-Care Activities

(CTFPHC) Canadian Task Force on Preventive Health Care

(CTT) Teoría clásica de los tests

(CVRS) Calidad de vida relacionada con la salud

(DASH) Dietary Approaches to Stop Hypertension

(DGC) Dispositivo de glucemia capilar

(DIF) Funcionamiento diferencial de los ítems

(DKAT) Diabetes Knowledge Assessment Test

(DKB) Diabetes Knowledge and Behaviour Questionnaire

(DKN) Diabetes Knowledge Assessment Scale

(DKQ) Diabetes Knowledge Questionnaire

(DKT) Diabetes Knowledge Test

(DM) Diabetes Mellitus

- (DMK) Diabetes Mellitus Knowledge Questionnaire
- (DPKT) Diabetes Patient Knowledge Test
- (DRNK) Diabetes-related Nutrition Knowledge
- (DS) Defensividad Sensorial
- (ECAs) Ensayos clínicos aleatorizados
- (EFA) Análisis Factorial Exploratorio
- (EFC) Análisis Factorial Confirmatorio
- (EMA) European Medicines Agency
- (EMPRO) Evaluating Measures of Patient Reported Outcomes
- (EMPRO) Evaluating the Measurement of Patient-Reported Outcomes
- (EORTC QLQC15-PAL) European Organization for Research and Treatment of Cancer Quality of Life 15 items Questionnaire for Palliative Care
- (EORTC QLQ-C30) European Organization for Research and Treatment of Cancer Core Quality of Life Questionnaire 30 items
- (EORTC QLQ-FA12) European Organization for Research and Treatment of Cancer Quality of Life 12 items Questionnaire
- (EORTC QLQ-LMC21) European Organization for Research and Treatment of Cancer Colorectal Liver Metastases- 21items
- (EQ-5D) EuroQol-5D
- (ESAS) Edmonton Symptom Assessment System
- (ESAS: AM) Edmonton Symptom Assessment System: Ascites Modification
- (ESAS:AM-24h) Edmonton Symptom Assessment System: Ascites Modification - 24 hours

(FACIT-Pal) Functional Assessment of Chronic Illness Therapy- Palliative Care

(FACT-F) Functional Assessment of Cancer Therapy-Fatigue

(FACT-G) Functional Assessment of Cancer Therapy- General

(FDA) Food and Drug Administration

(FFQ) Food Frequency Questionnaire

(FID) Federación internacional de documentación

(FIFS) Four-Item Fatigue Scale

(FRC) Fatiga relacionada con el cáncer

(GRADE) Grading of Recommendations, Assessment, Development and Evaluation/Calificación de recomendaciones, valoración, desarrollo y evaluaciones

(GRRAS) Guidelines for Reporting Reliability and Agreement Studies

(HADS) Hospital Anxiety and Depression Scale

(HbA1c) Hemoglobina glicosilada

(HQLI) Hospice Quality of Life Index

(HQLS) Hospice Quality of Life Scale

(HRQL) health-related quality of life

(ICC) Cociente de correlación intraclase

(ICD-10) International Classification of Diseases-Tenth Revision

(IMC) Índice de masa corporal

(IR) Índice de representatividad

(IRT) Teoría de respuesta al ítem

(ISOQOL) International Society for Quality of Life Research

(ISPOR) International Society for Pharmacoeconomics and Outcomes Research

(IVC) Índice de validez de contenido

(KMO) Kaiser Meyer Olkin

(K-MQLS) Korean- McMaster Quality of Life Scale

(LoA) Límites de acuerdo

(LVQOL) Low Vision Quality of Life

(MCG) Monitorización continua de glucosa

(MDDM2) Motiv.Diaf-DM2 Questionnaire

(MFI-20) Multidimensional Fatigue Inventory- 20 items

(MGCFA) Análisis factorial confirmatorio multigrupo

(MIC) Cambio mínimo importante

(ML) Estimación de máxima verosimilitud

(MMS) Mini Mental State

(MQLS) McMaster Quality of Life Scale

(NANDA) North American Nursing Diagnosis Association

(NHMRC) National Health and Medical Research Council

(NHP) Nottingham Health Profile

(NICE) National Institute for Health and Clinical Excellence

(OMS) Organización Mundial de la Salud

(PCCRP) Practicing Chiropractors' Committee on Radiology Protocols

(PDAQ) Perceived Dietary Adherence Questionnaire

(PDQ) Personal Diabetes Questionnaire

(PNPC) Palliative Care for Cancer Patients

(PQ) Perform Questionnaire

(PQLI) Palliative Care Quality of Life Instrument

(PRESS) Peer Review of Electronic Search Strategies / Revisión por Pares de las Estrategias de Búsqueda Electrónica

(PRISMA) Preferred Reporting Items for Systematic Reviews and Meta-Analyses / Elementos de Informe Preferidos para Declaraciones Sistemáticas y Metanálisis

(PRISMA-P) Directrices de Elementos de Informe Preferidos para Revisiones Sistemáticas y Protocolos de Metanálisis

(PRO) Resultados informados por el paciente

(PROM) Medidas de resultado informadas por el paciente

(PROMIS) Patient-Reported Outcomes Measurement Information System

(PROSPERO) protocol of this systematic review was registered in the International Prospective Register of Systematic Reviews / Registro Prospectivo Internacional de Revisiones Sistemáticas

(PSAA) Perfil sensorial del adolescente/adulto

(PSI) Índice de separación de personas

(QAREL) Quality Appraisal of Diagnostic Reliability

(QLI) Spitzer Quality of Life Index

(QUADAS) Quality Assessment of Diagnostic Accuracy Studies

(RDKS) Revised Diabetes Knowledge Scale

(RMSEA) Error cuadrático medio de aproximación

- (ROC) Característica operativa del receptor
- (SAC-MOT) Scientific Advisory Committee of the Medical Outcomes Trust herramienta
- (SAS) Servicio Andaluz de Salud
- (SAS) Symptom Assessment Scale
- (SCQ) Social Communication Questionnaire
- (SDC) Cambio mínimo detectable
- (SDKS) Simplified Diabetes Knowledge Scale
- (SDKT) Simplified Diabetes Knowledge Test
- (SE (θ)) Error estándar de theta
- (SEM) Error estándar de medición
- (SF-36) Short Form-36 Health Survey
- (SIGN) Scottish Intercollegiate Guidelines Network
- (SL) Mínimos cuadrados
- (SORT) Strength of Recommendation Taxonomy
- (SP) Sensory Profile
- (SRMR) Raíz cuadrada media residual estandarizada
- (SSP) Short Sensory Profile
- (SSP-2-PL) Polish version of the Short Sensory Profile, 2nd edition
- (STARD) STAndards for the Reporting of Diagnostic accuracy studies
- (STROBE) STrengthening the Reporting of Observational Studies in Epidemiology

(TLI) Índice de Tucker- Lewis

(UKDDQ) United Kingdom Diabetes and Diet Questionnaire

(USPSTF) U.S. Preventive Services Task Force

(VAS) visual analogue scale

(VCM1) Vision Related Quality of Life Core Measure

(VF-14) Visual Functioning questionnaire

Validación de herramientas de evaluación en terapia ocupacional y enfermería a través de la metodología COSMIN

RESUMEN

La lista de verificación de los Estándares basados en el COnsenso para la selección de INstrumentos de Medición de la salud (COSMIN) es una herramienta desarrollada para evaluar la calidad metodológica de estudios individuales sobre las propiedades de medición de los cuestionarios de evaluación, así como para la validación de dichos cuestionarios.

En unos 10 puntos, la guía COSMIN describe la metodología de revisiones sistemáticas sobre cuestionarios existentes y detalla unos patrones para cuestionarios de evaluación en ciencias de la salud, con el fin de estandarizar el proceso de validación siguiendo una misma metodología y evaluando cada una de las propiedades psicométricas recogidas en su guía.

El objetivo general fue conocer y aplicar de una forma precisa la metodología COSMIN para la realización de revisiones sistemáticas sobre herramientas de evaluación y para la validación de cuestionarios.

Se llevaron a cabo 3 revisiones sistemáticas, en las cuales se evaluó la calidad metodológica y la calidad de los resultados para cada estudio sobre una propiedad de medición utilizando la COSMIN Risk of Bias checklist, los criterios publicados por Terwee y Prinsen y el nivel de evidencia general de cada herramienta por propiedad psicométrica mediante el método GRADE.

Además de ello, se validó la escala Sensory Profile para población española siguiendo la metodología COSMIN.

Resultados

Para la primera revisión sistemática, "Nutritional status measurement instrument: a systematic psychometric review", el cuestionario Food Frequency Questionnaire (FFQ) fue el recomendado para evaluar la ingesta dietética en personas con diabetes.

Para la segunda revisión sistemática, "The diabetes knowledge measurement instrument: a systematic psychometric review", los cuestionarios Michigan Diabetes Knowledge Test (DKT), Diabetes Knowledge Questionnaire (DKQ) y Personal Diabetes Questionnaire (PDQ) fueron los instrumentos seleccionados para medir el conocimiento dietético en personas con diabetes tipo 1 y 2.

Para la tercera revisión sistemática, "Cancer-related fatigue measures in palliative care: a psychometric systematic review", Edmonton Symptom Assessment System (ESAS), Problems and Needs in Palliative Care Questionnaire (PNPC), European Organization for Research and Treatment of Cancer Quality of Life 15 items Questionnaire for Palliative Care (EORTC QLQ-C15-PAL) y Palliative Care Quality of Life Instrument (PQLI) fueron las herramientas recomendadas para evaluar la fatiga en pacientes con cáncer en tratamiento paliativo.

Para el estudio de validación del cuestionario AASP "Validation of the Adolescent/Adult Sensory Profile (AASP) for Spain", los resultados obtenidos mostraron un α de Cronbach de 0.83 para todos los ítems y un ICC para cada uno de los 4 cuadrantes de 0.90, 0.90, 0.90 y 0.83. Se compararon los ítems de la versión original de AASP-60 con los ítems de la nueva versión adaptada al español de AASP-42, y los niveles de correlación encontrados fueron altos ($r = 0.88-0.97$; $p < 0.01$).

Conclusiones

Es necesario tener un conocimiento exhaustivo sobre la metodología COSMIN y su aplicación para la realización de revisiones sistemáticas, con el fin de posibilitar la selección del instrumento más apropiado y facilitar la validación de este tipo de herramientas. Además de ello, el conocimiento de dicha metodología nos proporcionaría las herramientas necesarias para la validación de escalas para una población clínica en concreto.

Palabras clave: psicometría; sector de la salud; bioestadística; métodos estadísticos en psicología.

Validation of measurement instruments in occupational therapy and nursing through the COSMIN methodology

ABSTRACT

The CONsensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) checklist is a tool developed to assess the methodological quality of individual studies on the measurement properties of assessment questionnaires and the validation of these questionnaires.

In about ten points, the COSMIN guide describes the methodology of systematic reviews on existing questionnaires and details some patterns for evaluation questionnaires in health sciences to standardize the validation process following the same methodology and evaluating each of the psychometric properties included in your guide.

The general objective was to know and apply the COSMIN methodology precisely to carry out systematic checks on evaluation tools and for the validation of questionnaires.

Three systematic reviews included in the Doctoral Thesis were carried out in which the methodological quality and the quality of the results were evaluated for each study on a measurement property using the COSMIN Risk of Bias checklist and the criteria published by Terwee and Prinsen, and the general level of evidence of each tool by psychometric property following the GRADE method. In addition to this, the Sensory Profile scale was validated for the Spanish population following the COSMIN methodology.

Results

For the first systematic review, "Nutritional status measurement instrument: a systematic psychometric review," the Food Frequency Questionnaire (FFQ) was recommended to assess dietary intake in people with diabetes.

For the second systematic review, "The diabetes knowledge measurement instrument: a systematic psychometric review," the Michigan Diabetes

Knowledge Test (DKT), Diabetes Knowledge Questionnaire (DKQ), and Personal Diabetes Questionnaire (PDQ) were the instruments selected to measure dietary knowledge in people with type 1 and 2 diabetes.

For the third systematic review, "Measures of cancer-related fatigue in palliative care: a systematic psychometric review," Edmonton Symptom Assessment System (ESAS), Problems and Needs in Palliative Care Questionnaire (PNPC), European Organization for Research and Treatment of Cancer Quality of Life 15 items Questionnaire for Palliative Care (EORTC QLQ-C15-PAL) and Palliative Care Quality of Life Instrument (PQLI) were the recommended tools to assess fatigue in cancer patients undergoing palliative treatment.

For the validation study of the AASP questionnaire "Validation of the Adolescent/Adult Sensory Profile (AASP) for Spain," the results obtained showed a Cronbach's α of 0.83 for all items and an ICC for each of the four quadrants of 0.90, 0.90, 0.90 and 0.83. The items of the original version of the AASP-60 were compared with the items of the new version adapted to Spanish of the AASP-42, and the levels of coincidences found were high ($r = 0.88-0.97$; $p < 0.01$).

Conclusions

It is necessary to have an exhaustive knowledge of the COSMIN methodology and its application to carry out systematic checks to enable the selection of the most appropriate instrument and to facilitate the validation of this type of tool. In addition to this, knowledge of said methodology would provide us with the necessary tools to validate scales for a specific clinical population.

Keywords: psychometry; health sector; biostatistics; statistical methods in psychology.

I. INTRODUCCIÓN

I. INTRODUCCIÓN

Este trabajo de investigación consiste en la validación, análisis y elección de herramientas de evaluación en el área de enfermería y terapia ocupacional mediante el uso de la metodología COSMIN.

Comienza con la presentación del marco teórico, que consta de 8 puntos en los que está fundamentado cada uno de los estudios que se presentan en la tesis doctoral:

En el punto 1 se realiza una descripción detallada sobre cuestionarios de evaluación en ciencias de la salud (definición, principios y propiedades; desarrollo de un cuestionario, validez y adaptación, y causas que pueden alterar tanto la validez como la fiabilidad de estos).

En el punto 2 se efectúa una descripción y definición de todas las propiedades psicométricas, ofreciendo información sobre los índices estadísticos más idóneos para cada una de ellas, así como sus valores de referencia para el posterior análisis de los resultados.

En el punto 3 se lleva a cabo una amplia descripción sobre la normativa COSMIN, definiendo y desarrollando los 10 puntos de la guía COSMIN para revisiones sistemáticas.

En el punto 4 se describen los criterios de Terwee y Prinsen para valorar la calidad de los resultados, detallando cuáles son sus criterios de puntuación y valores de referencia.

El punto 5 se centra en el método GRADE modificado para la evaluación de revisiones sistemáticas enfocadas a escalas de valoración, puntualizando cuáles son sus cuatro niveles de evidencia y los factores que la aumenta o disminuyen.

Tanto en el punto 6 como el 7 se describen dos listas de verificación creadas para evaluar la calidad de estudios de tipo observacional: STARD y STROBE. La lista de verificación STARD fue la utilizada para evaluar la calidad de los trabajos de investigación incluidos en el artículo 1: "Nutritional status measurement instrument: a systematic psychometric review", mientras que la lista de

verificación STROBE fue la empleada para el artículo 3: "Cancer-related fatigue measures in palliative care: a psychometric systematic review".

Para finalizar, en el punto 8 se realiza un compendio de las distintas poblaciones incluidas, tanto para el trabajo de campo (disfunción del procesamiento sensorial) como para las 3 revisiones sistemáticas (diabetes y cáncer).

A continuación, en el bloque IV se detallan los objetivos correspondientes a cada uno de los estudios de investigación incluidos para la elaboración de dicha tesis doctoral.

El siguiente bloque, marco empírico, muestra 3 revisiones sistemáticas efectuadas como parte del proceso inicial de análisis de resultados de las propiedades psicométricas, siguiendo la metodología COSMIN para diferentes herramientas de evaluación dirigidas a población con diabetes y cáncer.

En primer lugar, se llevó a cabo una selección de aquellos estudios que utilizaron herramientas dirigidas a evaluar el estado nutricional en diabetes ("Nutritional status measurement instrument: a systematic psychometric review", publicado en *International Journal of Environmental Research and Public Health* el 7 de agosto de 2020), el conocimiento dietético en diabetes ("The diabetes knowledge measurement instrument: a systematic psychometric review", publicado en *Journal of Advanced Nursing* el 9 de febrero de 2021) y la fatiga relacionada con el cáncer en cuidados paliativos ("Cancer-related fatigue measures in palliative care: a psychometric systematic review", dicho artículo se encuentra en proceso de revisión en la revista *European Journal of Cancer Care*). Se analizó la calidad metodológica para cada estudio sobre una propiedad de medición utilizando la COSMIN Risk of Bias checklist, la calidad de los resultados de estos aplicando los criterios publicados por Terwee y Prinsen y el nivel de evidencia general de cada herramienta por propiedad psicométrica siguiendo el método GRADE.

Dentro del marco empírico, también se incluyó la ejecución del estudio de las propiedades psicométricas de la versión española del cuestionario Adolescent/Adult Sensory Profile (AASP) en "Validation of the adolescent/adult sensory profile (AASP) for Spain" (dicho artículo se encuentra en proceso de revisión en la revista *Australian Occupational Therapy*), mostrándose la

metodología empleada y los resultados obtenidos para dicho cuestionario en la población española.

En el último bloque se exponen las deliberaciones sobre la discusión general, principales limitaciones, implicaciones para la práctica clínica y recomendaciones para futuras líneas de investigación. Se continúa con las conclusiones, para terminar con los anexos que pertenecen a documentación de interés.

II. JUSTIFICACIÓN

II. JUSTIFICACIÓN

Es habitual, en ciencias de la salud, tanto en la práctica clínica como en investigación emplear cuestionarios autoadministrados como una forma de medir constructos o variables que no se pueden observar. Estos son imprescindibles, ya que ofrecen una medida concluyente sobre resultados en salud centrados tanto en la persona como en su patología, y manifiestan la necesidad de incluir al paciente en la evaluación de sus resultados de salud (1-3). Dichos cuestionarios deben contemplar una serie de propiedades técnicas, entre ellas validez y fiabilidad, que nos ofrezcan garantías para corroborar que se está midiendo aquello que se quiere medir con un error de medición mínimo.

Estas herramientas de evaluación van a ayudar a identificar aspectos relevantes referentes a problemas de salud presentes en el individuo, como aquellos relacionados con el funcionamiento físico (cuidado e higiene personal, andar, cargar peso, subir escaleras, esfuerzos moderados o intensos), emocional (sufrimiento psicológico, falta de bienestar emocional, ansiedad y depresión), social (problemas de salud física o emocional que interfieren en la vida social, en la familia y con los amigos), de rol funcional, cognitivo (delirio, amnesia, demencia, alzhéimer, párkinson, accidente cerebrovascular, traumatismo craneal), de percepción de la salud general y el bienestar (salud actual, perspectivas y resistencia a enfermar) y en relación con síntomas específicos para cada enfermedad.

En ausencia de evidencia empírica y en respuesta al problema que supone seleccionar la herramienta de evaluación más adecuada, el comité directivo de COSMIN desarrolló una metodología para realizar revisiones sistemáticas con el objetivo de medir un constructo de interés en una población de estudio específica, permitiendo evaluar la calidad metodológica, la calidad de la evidencia y la calidad de los resultados (4,5). Este tipo de revisiones sistemáticas pueden proporcionar una descripción general completa de las propiedades de medición de estos cuestionarios, determinar el riesgo de sesgo de los resultados de los estudios individuales incluidos en dicha revisión y respaldar las recomendaciones

basadas en la evidencia para la selección del instrumento más adecuado para un propósito determinado (investigación o práctica clínica, aplicaciones discriminativas, evaluativas o predictivas) (6).

En unos 10 puntos, la guía COSMIN describe la metodología de revisiones sistemáticas sobre cuestionarios existentes, para los cuales se dispone de al menos alguna información sobre sus propiedades de medición, su uso con fines evaluativos y su selección para un propósito específico. Estos pasos a seguir incluyen: la búsqueda y selección sistemática de artículos relevantes, la evaluación de la calidad de los estudios elegibles, la medición de las propiedades psicométricas de los instrumentos incluidos, la valoración de aspectos de interpretabilidad y factibilidad, la formulación de conclusiones y recomendaciones, y su publicación (5).

Además de la metodología COSMIN (4,7,8) para evaluar la calidad de estos instrumentos de evaluación existen otras guías, como los estándares de American Psychological Association (APA) (9,10), Food and Drug Administration (FDA) (2,11), Agency for Healthcare Research and Quality (AHRQ) (12), Scientific Advisory Committee of the Medical Outcomes Trust (SAC-MOT) (6,13), la herramienta Evaluating the Measurement of Patient-Reported Outcomes (EMPRO) (basada en SAC-MOT) (14), los criterios de Terwee y Prinsen (5,7), y una lista de verificación publicada en 2016 por Francis et al (15). A pesar de existir otros estándares con dicho fin, la mayoría de estos métodos son mucho más breves y en sus pautas a seguir no describen ni explican completamente todos los pasos necesarios para determinar la calidad de este tipo de instrumentos de evaluación de manera sistemática; en sus directrices no se establece que la conclusión deba basarse en una consideración sistemática de toda la evidencia existente. Del mismo modo, muchas de estas pautas combinan estándares para la calidad de los estudios con criterios para calificar los resultados de las propiedades de medición en un solo estándar, y estos estándares no están lo suficientemente detallados para proporcionar calificaciones transparentes y sistemáticas del riesgo de sesgo de los estudios sobre dichas propiedades de medición.

Por todo esto, es necesario tener un conocimiento exhaustivo sobre la metodología COSMIN y su aplicación para la realización de revisiones

sistemáticas sobre un constructo de interés en una población de estudio específica, con el fin de posibilitar la selección del instrumento más apropiado y facilitar la validación de este tipo de herramientas, ya que con el instrumento idóneo se puede dar respuesta a un problema de salud aportando a los profesionales del ámbito sanitario un recurso ineludible de gran aplicación clínica que ayude a favorecer el proceso de valoración (recogida y recopilación de datos de la persona, familia y entorno), diagnóstico, planificación de estrategias (para prevenir, minimizar o corregir los problemas), tipos de intervención y su evaluación, con los beneficios que ello reporta en cuanto a salud, recursos humanos y materiales.

COSMIN también es importante porque presenta unos patrones para cuestionarios de evaluación en ciencias de la salud con el objetivo de estandarizar el proceso de validación de dichos cuestionarios siguiendo una misma metodología y evaluando cada una de las propiedades psicométricas recogidas en su guía (4).

Adicionalmente, esta tesis doctoral puede servir a investigadores como una guía en español sobre la metodología COSMIN (podría tratarse de la primera interpretación en español de la normativa COSMIN), para proporcionarles unos conocimientos sobre los pasos a seguir, guiarlos en la aplicación de dicho método, ofrecerles la posibilidad de mejorar la calidad de sus revisiones sistemáticas, ayudarles en la obtención de resultados más fiables y a favorecer la aplicación de un adecuado rigor metodológico a sus estudios. Así mismo, el hecho de que la guía COSMIN sea un método de evaluación suficientemente conocido y utilizado internacionalmente favorecerá la consiguiente difusión de los resultados obtenidos a la comunidad científica internacional, aportando una mayor visibilidad a la investigación científica española.

III. MARCO TEÓRICO

Cuestionarios de evaluación en
ciencias de la salud

1. CUESTIONARIOS DE EVALUACIÓN EN CIENCIAS DE LA SALUD

1.1 DEFINICIÓN DE CUESTIONARIO

Un cuestionario es una herramienta usada para recoger, universalizar y cuantificar la información. Su propósito es hacer comparable la información y estandarizar el procedimiento de la entrevista (16,17).

El uso de los cuestionarios es cada vez más habitual, tanto en el campo de la investigación como en el de la asistencia (práctica clínica). Se trata, junto con la entrevista, del método de recolección de datos más utilizado en la investigación, porque es más económico, alcanza a un número más amplio de individuos y posibilita el análisis, aunque no está exento de otras muchas limitaciones que pueden mermar el desarrollo de la investigación (17–20).

Generalmente, al hacer referencia a un cuestionario se habla de las escalas de evaluación. Estas escalas son cuestionarios o instrumentos que posibilitan un ascenso acumulativo de sus ítems, proporcionando puntajes globales al término de la evaluación. Y es en parte ese carácter acumulativo que los caracteriza, el que los diferencia de los inventarios de síntomas, los cuestionarios de recogida de datos, los formularios o las entrevistas estandarizadas. Por lo que, la mayoría de las veces, al referirnos a los cuestionarios, lo que estamos hablando es de medidas de resultado (21).

Al igual que las entrevistas, los cuestionarios fundamentan su información en sentimientos, en la validez que se le da a la información verbal con relación a percepciones, conductas o actitudes que informa el encuestado o datos difíciles de contrastar y transformar a un sistema de medida (a una puntuación). Es esta cualidad la que hace tan complicado instaurar en este tipo de instrumentos unos criterios de calidad (17).

1.2 PRINCIPIOS Y PROPIEDADES DE UN INSTRUMENTO DE MEDIDA

Construir y validar un cuestionario o escala de medida es un proceso relativamente complicado y precisa de una comprensión teórica sobre aquello que se pretende medir, así como avanzados conocimientos sobre estadística y determinados programas informáticos para su realización. Con ello, se intentará conseguir que el nuevo cuestionario responda con adecuados criterios de fiabilidad y validez. Cualquier instrumento de medida debe de cumplir las siguientes características (16):

- 1) Ser viable, sencillo y aceptado por usuarios pacientes e investigadores (reportar viabilidad).
- 2) Ser preciso y fiable, es decir, mostrar el menor número de errores posibles en la medición (fiabilidad).
- 3) Ser apropiado para el problema que se intenta medir (validez de contenido).
- 4) Manifestar la teoría subyacente en el concepto o fenómeno que se pretende medir (validez de constructo).
- 5) Ser idóneo para medir cambios, tanto en la respuesta de un mismo individuo como en diferentes individuos a través del tiempo (sensibilidad al cambio).

1.3 PROCESO PARA EL DESARROLLO DE UN CUESTIONARIO

Para la correcta elaboración de un cuestionario debemos ejecutar los siguientes pasos:

1.3.1 Definición del constructo o aspecto que pretendemos medir

Para poder *definir el constructo* se debe de tener una idea muy clara de aquello que se quiere medir. Este primer paso puede exigir tanto una revisión de la bibliografía como la consulta a un grupo de expertos en la materia. Sean conductas, actitudes o conocimientos, se tiene que concretar de una forma clara y concisa cuál es el objeto de la medida y establecer, para así poder conocer, aquellas teorías que respaldan la definición acordada. Un mismo problema puede

explicarse desde diferentes perspectivas teóricas y, por consiguiente, pueden exponerse definiciones desiguales de un mismo constructo. La validez de constructo determinará el grado en que el cuestionario muestra las teorías relevantes del fenómeno que evalúa (20).

1.3.2 Finalidad del cuestionario

Consiste en incluir el contenido del cuestionario, mostrar una definición sobre la población a la que va dirigido, y exponer tanto la manera en la que se va a administrar como el formato en el que se va a presentar (22). El propósito del cuestionario va a establecer en gran parte el contenido de sus ítems, diversos aspectos que guardan relación con el estudio de la recogida de datos y su estructura.

Cuando comenzamos con el desarrollo de un cuestionario, debemos o bien tener en consideración a la población a la que se va a administrar, o bien conocer en qué población (patologías, edad, intereses, nivel cultural...) se ha validado (en caso de que se elija un cuestionario ya validado) (22).

Por otra parte, en función del tema elegido para su estudio, deberemos determinar cómo va a ser administrado este: si será autoadministrado, si la recogida de información será vía telefónica o si se formalizará mediante entrevista personal. Es importante tener en cuenta todos estos aspectos, puesto que condicionarán el modo de redactar las preguntas o incluso le darán un formato heterogéneo al cuestionario (19).

Las formas de recogida de información citadas previamente pueden tener sus ventajas e inconvenientes. Las entrevistas personales o las encuestas telefónicas habitualmente tienen una tasa de respuesta más favorable que aquellas que se envían por correo postal, pero tienen un coste más alto. Además, si en la encuesta telefónica el cuestionario es demasiado extenso, puede terminar cansando al participante. En ambos casos, es necesario entrenar al entrevistador, y en el caso de los cuestionarios autoadministrados, adjuntar un mínimo de instrucciones (19).

1.3.3 Elaboración de los ítems

Los cuestionarios están formados por un número determinado de ítems. El ítem se define como la unidad fundamental de información de un cuestionario de evaluación, y habitualmente se compone de una pregunta y una respuesta cerrada (17).

1.3.4 Número de ítems

Generalmente se estima que el número mínimo adecuado de ítems para el aspecto a medir estaría establecido en 6, pero podría variar desde 10 a 90, de modo que se puedan englobar de una forma proporcionada todas las dimensiones concretadas *a priori* en el constructo. Se aconseja efectuar el doble de ítems de los que serán necesarios para una versión definitiva del instrumento (22).

1.3.5 Contenido

Los cuestionarios pueden ser unidimensionales o multidimensionales en función de su contenido. Para los primeros, aproximadamente más del 80% de los ítems evalúan una misma dimensión, mientras que para los segundos los ítems miden dos o más dimensiones (17).

1.3.6 Definición y disposición

La definición de cada ítem debe ser mutuamente excluyente y exhaustiva. Del mismo modo, al enunciar la pregunta debemos de considerar factores como la aceptabilidad por parte del sujeto que recibe la pregunta y su comprensión (importante adecuar el tipo de elección de respuestas y el lenguaje al nivel sociocultural de los participantes). Para ello, encontramos una serie de directrices a la hora de redactar las preguntas, como no utilizar palabras que induzcan a una reacción redundante, emplear preguntas que sean breves y fáciles de entender, eludir el uso de la interrogación “por qué”, no formular preguntas en forma negativa, prescindir de preguntas que exijan realizar esfuerzos de memoria o cálculos y nunca enunciar preguntas en las que una de las opciones de respuesta sea tan tentadora que apenas pueda ser rechazada (22).

Tras haber redactado cada uno de los ítems, estos deben ser ordenados. En muchos casos, el propio cuestionario servirá de guía sobre el orden establecido, pero si no fuera el caso, se pueden disponer de un modo aleatorio.

1.3.7 Identificación de los sesgos al completar el cuestionario

También es probable identificar los posibles sesgos a la hora de completar el cuestionario a través del diseño de los ítems y, con ello, perfeccionar la validez del instrumento. Aunque existen un gran número de sesgos, algunos de los más comunes son los siguientes: el *error de tendencia central* (consiste en elegir entre las distintas opciones de respuesta, las centrales), la *deseabilidad social* (la respuesta va en concordancia con lo socialmente aceptable), el *sesgo de aprendizaje* o de *proximidad* (incita a responder de manera análoga a las respuestas anteriores) y el *error lógico* (se origina cuando el participante piensa que todos los ítems que están relacionados reciben idéntica puntuación) (23).

Por último, debemos de tener siempre en cuenta como está redactado el cuestionario, el formato y el orden de las preguntas o de las categorías de respuesta, ya que esto podría inducir a sesgos en el momento de recoger la información.

1.3.8 Clasificación de las respuestas

Según el tipo de respuestas o número de opciones posibles, estas pueden ser: dicotómicas (dos únicas respuestas), policotómicas (tres o más alternativas de respuesta) o analógicas (diferentes en función del sistema analógico utilizado: lineal, numérica, verbal, etc) (22).

1.3.9 Puntuación de los ítems

Es imprescindible definir el sistema de puntuación que se va a emplear, este puede ser simple o ponderado. Hablamos de ítems simples cuando la puntuación directa se consigue con el total de la suma de respuestas acertadas o de los valores facilitados en cada opción, mientras que los ítems ponderados se dan cuando el

total de cada opción de respuesta no es igual o no se concede el mismo valor a cada uno de los aciertos (17).

1.3.10 Fase de validación

Una vez concretada la información y expuestas y definidas el número de preguntas que serán incluidas y ordenadas en el cuestionario, procede efectuar una prueba piloto y evaluar las propiedades psicométricas del cuestionario (18,20,24,25).

1.3.11 Prueba piloto o pre-test

Se administra un borrador del cuestionario a unos 5 u 8 participantes, siendo conveniente que estos se asemejen a los individuos de la muestra. Esta prueba piloto permitirá distinguir los tipos de preguntas más apropiados, revelará si es acertada la categorización de las respuestas, si es lógico el ordenamiento interno, si la duración es aceptable para los participantes, si el enunciado es comprensible y adecuado, si las preguntas poseen la extensión conveniente y si existe intransigencia psicológica o rechazo a alguna de las preguntas (8,26).

Para la realización de la prueba piloto, los métodos empleados son: entrevistas informales, encuestas sobre comprensión de las preguntas, valoración del cuestionario por parte de los participantes en el estudio y grupos focales de la población diana (17).

1.3.12 Evaluación de las propiedades psicométricas del cuestionario

Al diseñar un cuestionario de medición que proporciona una puntuación sobre un aspecto de la salud y permite establecer comparaciones en un mismo individuo o en otros individuos en diferentes momentos, este debe ser válido y fiable (27–29).

La **validez** establece el grado en el que un instrumento mide aquello para lo que ha sido diseñado. Se trata de una pieza indispensable, tanto para diseñar un cuestionario como para confirmar la utilidad de la medida efectuada. Puede

calcularse de múltiples formas, a través de la validez lógica o aparente, de la validez de contenido, de la validez de criterio y de la validez de constructo (18). En el caso de los cuestionarios utilizados para pruebas diagnósticas (por ejemplo, un test para diagnosticar malnutrición) encontramos dos conceptos más: la sensibilidad (facultad que posee un cuestionario para encontrar cambios en los sujetos o atributos evaluados, tras una intervención) y la especificidad (posibilidad de catalogar adecuadamente a un sujeto sano como tal). Esto significa que, cuanto más sensible y específico sea un cuestionario, menos falsos positivos y falsos negativos contendrá. Existen varios tipos de validez en relación con un instrumento que precisan de métodos estadísticos de mayor o menor complejidad (20).

- *Validez lógica o aparente*

Es el grado en el que un instrumento mide lo que pretende medir a juicio tanto de los propios sujetos como de los expertos. La determinación sobre si las preguntas han de tener o no validez lógica debe tomarse anteriormente al comienzo de su redacción. En el caso de que dichas cuestiones carezcan de validez lógica es muy presumible que los individuos estudiados rehúsen contestar a las preguntas. Solamente en ciertos casos es conveniente la formulación de preguntas escasas de validez lógica (20).

- *Validez de contenido*

Se muestra cuando los ítems que conforman parte del cuestionario son apropiados para medir el constructo que se pretende medir, y se consigue mediante el consenso de un panel de expertos que valoran la idoneidad de los ítems del instrumento con el fin de medir aquello que se pretende.

Los métodos empleados para dicha medición son: el método Delphi, el modelo Fehring, la estimación de magnitud y la metodología Q; y los métodos estadísticos habitualmente utilizados son: el índice de validez de contenido (IVC), la prueba no paramétrica (W) de Kendall y la prueba de V de Aiken, encargados todos ellos de evaluar el grado de correspondencia entre el grupo de expertos y la validez de contenido que presentan los ítems con relación al cuestionario (30).

- *Validez de criterio*

Esta propiedad psicométrica hace posible reconocer si nuestro cuestionario mide lo que en teoría dice medir o si está midiendo cosas distintas a las que se pretendía. La forma más común es utilizar un instrumento que ya se encuentre estandarizado a modo de “gold standard” y comparar los resultados alcanzados entre ambos cuestionarios. Si existe equivalencia, significa que nuestro cuestionario es capaz de evaluar lo mismo (30).

- *Validez de constructo*

Es la propiedad más compleja de evaluar. Se emplea para distinguir si el instrumento se adapta a un modelo predictivo estadístico con relación al constructo que se pretende medir, utilizando para ello tanto un análisis factorial exploratorio (EFA) como un análisis factorial confirmatorio (CFA), siendo este un método de análisis estadístico de tipo multivariante que minimiza un grupo de variables interrelacionadas en un número de variables comunes, y estas reciben el nombre de “factores” o “dimensiones”. Esto origina que muchos de estos instrumentos dispongan de más de un “factor” o “dimensión”, dentro del cual se asocian una serie de ítems que evalúan ese mismo factor (30,31).

La **fiabilidad** hace referencia a la estabilidad en el tiempo y a la precisión de la medida. Si a un conjunto de individuos le pasamos un mismo cuestionario dos o más veces, posibilitando entre ambas pruebas un determinado intervalo de tiempo, podremos hablar de fiabilidad siempre que los sujetos sean proclives a conservar la misma puntuación o queden ordenados de forma similar en las diferentes aplicaciones (20,31).

Existen diversos métodos estadísticos para medir la fiabilidad de un instrumento: la consistencia interna (cuando las puntuaciones de los individuos evaluados no se modifican tras realizar una repetición empleando para una misma herramienta distintos conjuntos de ítems), la fiabilidad test-retest (cuando existe una similitud en los resultados tras la medición de la misma muestra en dos momentos o situaciones diferentes), la fiabilidad inter-observador (para la misma ocasión por distintos profesionales) y la fiabilidad intra-observador (en diferentes ocasiones por los mismos profesionales).

- *Consistencia interna*

Explica el grado de correlación que existe entre cada uno de los diferentes ítems que forman el instrumento de evaluación. Su fin es confirmar si los ítems que miden un mismo constructo son homogéneos entre sí y si los resultados son dependientes de los ítems que hemos administrado (32,33). El más utilizado es el estadístico de correlación alfa de Cronbach (34), el cual siempre toma valores ≤ 1 , siendo aceptables ≥ 0.70 para cada escala unidimensional o cada subdimensión de la escala (33,35).

- *Fiabilidad test-retest*

El objetivo de la fiabilidad test-retest es la evaluación de la estabilidad temporal del instrumento de evaluación. Para ello, se efectúan dos evaluaciones a un mismo individuo a través de la administración del instrumento en similares condiciones, pero en momentos distintos (36), así se logra el grado de consistencia entre los puntajes de dos evaluaciones efectuadas por un mismo evaluador en dos momentos de tiempo diferentes (32).

- *Fiabilidad intra e inter-observador*

Consiste en el nivel de consistencia presente en las puntuaciones de un individuo que han sido concedidas por un mismo evaluador (intra-observador) o por evaluadores diferentes (inter-observador) (37). El objetivo de la fiabilidad intra-observador es valorar el nivel de concordancia del evaluador al calificar nuevamente la misma evaluación, equiparando los resultados consigo mismo, mientras que la finalidad de la fiabilidad inter-observador es saber qué rango de consistencia hay entre dos o más evaluadores cuando estos puntúan la evaluación de un mismo individuo.

1.4 CAUSAS QUE PUEDEN ALTERAR TANTO LA VALIDEZ COMO LA FIABILIDAD DE UN CUESTIONARIO

Existen una serie de factores que pueden alterar tanto la validez como la fiabilidad de un cuestionario: la improvisación a la hora de confeccionar el cuestionario, que no se halle validado en el marco en el que se va a aplicar

(utilizar cuestionarios diseñados para adultos en niños), dificultad para ser entendido o instrucciones complejas, insuficiente motivación del sujeto a la hora de responder, momentos o lugares inapropiados para administrar el cuestionario, contar con una muestra escasa o muy irregular, que se dé el fenómeno de la deseabilidad social al responder (dar las respuestas para ser aceptado socialmente o para agradar al investigador), etc (23).

1.5 VALIDEZ Y ADAPTACIÓN DE CUESTIONARIOS

El método de traducción y adaptación de un cuestionario es algo más complejo que la traducción de su lengua nativa a la lengua de destino, ya que es imprescindible garantizar que las puntuaciones alcanzadas con el instrumento ya traducido son equiparables a las conseguidas con el instrumento original. Para lograr dicha equivalencia debemos de tener en cuenta cuatro aspectos: el marco cultural donde se va a efectuar la adaptación, los aspectos técnicos del proceso de desarrollo y de la adaptación del cuestionario, la administración del mismo y la representación de las puntuaciones (17).

Propiedades psicométricas de
los cuestionarios

2. PROPIEDADES PSICOMÉTRICAS DE LOS CUESTIONARIOS

La **psicometría** es la disciplina encargada de abordar todas aquellas técnicas, métodos y teorías empleadas para la medición y cuantificación de las variables psicológicas de la mente humana. La psicometría abarca tanto la teoría como la construcción de test, pruebas y otras técnicas de medición válidas y fiables. Comprende la aplicación y la obtención de operaciones estadísticas que autorizan a determinar si un test o una prueba son o no válidas para medir una variable o una conducta psicológica anteriormente definida (38,39).

Los contenidos teóricos de la psicometría se sustentan básicamente en dos grandes grupos: teoría de los test (construir, validar y aplicar los test) y escalamiento (contiene los métodos para la obtención de escalas psicológicas y psicofísicas) (38,39).

Dentro del campo de la psicometría debemos de centrarnos en la rama de la **clinimetría**, que es la base donde se sustenta dicha tesis doctoral. El término clinimetría fue incluido por Alvan R. Feinstein en 1982 para hacer referencia a la construcción de una variedad de instrumentos empleados para medir o describir signos físicos, síntomas y/u otros fenómenos clínicos (40). Se describe como una rama elemental de la medicina que se ocupa de la especificación, identificación y medición de fenómenos clínicos en sujetos, abarcando los distintos elementos que conforman la relación entre el paciente y el profesional sanitario (41). Se trata de un proceso a través del cual, por medio de una serie de instrumentos de medición, valoramos una variedad dispar de fenómenos clínicos objetivos (los cuales disponen de una base empírica), permitiendo recoger la información precisa para llevar a cabo un diagnóstico correcto, evaluar la actividad de una patología, barajar acciones terapéuticas y especular con respecto a los pacientes sobre un determinado pronóstico (6,42).

A pesar de ello, tanto en clinimetría como en psicometría, las propiedades analizadas de los cuestionarios son similares, siendo estas: validez de contenido, validez estructural, consistencia interna, fiabilidad, error de medición, validez de

criterio convergente, validez de criterio, validez transcultural, prueba de hipótesis y sensibilidad al cambio.

2.1 VALIDEZ DE CONTENIDO

La validez de contenido es la propiedad psicométrica más importante y se define como el grado en que el contenido de un instrumento es un reflejo adecuado del constructo que se pretende medir (7,43). Debemos determinar que el criterio de los expertos es decisivo en la primera etapa, tanto para la construcción como para la validación de escalas en el campo de la salud.

En esta **etapa inicial** donde se lleva a cabo la validación, uno de los métodos empleados más frecuentes es el Delphi (44). Este está fundamentado en un proceso organizado minuciosamente para sintetizar y coleccionar el conocimiento de un conjunto de expertos a través de una sucesión de cuestionarios que se adjuntan con una serie de comentarios de opinión controlados (45). Fue instaurado por RAND Corporation, concretamente por Helmer y Dulkey, quienes implantaron en 1953 un sistema de repetición que incluía un proceso de realimentación controlada, que definieron como Delfos, en honor al famoso oráculo de la Grecia antigua, con el fin de disminuir los inconvenientes mostrados por otros métodos y aprovechar al máximo las ventajas de los métodos fundamentados en juicios de expertos (44). Para Delphi concurren numerosas publicaciones, la más destacada es una revisión sistemática de García y Suárez, en la cual se presta especial atención a una serie de conceptos, y en la que se describe su historia, ventajas, usos, características y se simplifica su elaboración dentro del campo de la salud, partiendo de la diferenciación en 3 fases: una primera fase de preparación (elección de expertos, preparación del instrumento y deliberación para determinar la vía de consulta), la fase de consulta (turno de consultas, proceso estadístico y realimentación) y, por último, la fase de aprobación e información de los resultados (44).

Este panel Delphi (3,6,46,47) establece que la validez de contenido debe evaluarse emitiendo un juicio sobre la relevancia y la amplitud de los elementos. La categoría de los ítems debe valorarse juzgando si estos son relevantes para el constructo que se va a medir, para la población de estudio y para el propósito del instrumento (8,48). Cuando se desarrolla una nueva herramienta, el enfoque y el

detalle del contenido del instrumento deben coincidir con la población objetivo, ya que si el instrumento se utiliza posteriormente en otra diferente que no sea la población objetivo original para la que se desarrolló, se debe evaluar si todos los elementos son relevantes para esta nueva población de estudio (8,48). Debido a que el enfoque está en el instrumento, los pacientes deben ser considerados expertos a la hora de juzgar la relevancia de los ítems; muchas observaciones faltantes sobre un ítem pueden ser una indicación de que el ítem no es relevante para la población o está formulado de manera ambigua.

Para evaluar la amplitud de los ítems se deben tener en cuenta tres aspectos: la cobertura de contenido de los ítems, la descripción de los dominios y la base teórica. Los dos primeros se refieren a la pregunta de si todos los aspectos relevantes del constructo están cubiertos por los ítems y los dominios, mientras que el último, la base o fundamento teórico, hace referencia a la disponibilidad de una descripción clara del constructo y la teoría en la que se basa (8,48).

El objetivo final al evaluar la validez de contenido de un instrumento es juzgar si este cumple con los criterios predefinidos para lo que constituye una buena validez de contenido. El estudio Delphi establece 10 criterios relacionados con 3 aspectos de dicha propiedad psicométrica en un instrumento: relevancia (todos los elementos que componen el instrumento deben ser relevantes para el constructo de interés dentro de una población y contexto de uso específico), exhaustividad (no debe faltar ningún aspecto clave del constructo) y comprensibilidad (los pacientes deben entender los elementos según lo previsto) (8). Estos 10 criterios se formulan como preguntas al evaluar la validez de contenido y estas preguntas deben responderse y resumirse en una calificación general de relevancia, exhaustividad, comprensibilidad y validez general de contenido para cada puntuación o subpuntuaje por separado, teniendo como referencia la información sobre cómo se desarrolló el instrumento, la evidencia de los estudios de validez de contenido y el contenido de la nueva herramienta (8).

En una **segunda etapa** se constituye un panel de expertos, un conjunto de especialistas independientes y acreditados por lo menos en uno de los campos a evaluar presentes en el programa, que se reúnen con el objetivo de emitir un juicio consensuado y colectivo sobre dicho programa. Según se les requiera, el juicio a emitir puede aludir a la ejecución práctica o a los efectos de la totalidad o

una parte del programa. El panel de expertos se forma principalmente para evaluar un conjunto de procedimientos estándar, continuando con un modo de trabajo determinado para reunirse y obtener un juicio (3,6,8,46,47). Los criterios que se aplican a la hora de seleccionar a cada uno de los expertos que van a constituir dicho panel son: experiencia profesional en el campo tratado como requisito indispensable, independencia del experto con relación al programa que se va a valorar y capacidad para trabajar en equipo. La elección de estos se puede realizar mediante una selección rápida (listado de expertos que son reconocidos en un campo concreto y para los que es requisito indispensable verificar su independencia en relación con el programa evaluado) o bien mediante una selección por etapas (conveniente definir en primer lugar el perfil de los expertos en función del tema que se va a tratar en la evaluación) (3,6,8,46,47).

En la **última etapa** se lleva a cabo una prueba piloto. Esta se considera parte de la validez de contenido (evaluación de la comprensibilidad). Se sugiere que en esta prueba intervengan entre 5 y 8 participantes de similares características a las de la población objetivo, controlando sus respuestas y el tiempo que emplean en darlas (8,26). La finalidad de dicha prueba es valorar peculiaridades de la nueva versión, el contexto de aplicación y/o hallar posibles debilidades o factores que puedan interferir en su adecuada ejecución (8).

Así, un ejemplo del proceso para la consecución de la medición de la propiedad psicométrica más importante, la validez de contenido, estaría reflejado en un artículo incluido en una de nuestras revisiones sistemáticas publicada, "The diabetes knowledge measurement instrument: a systematic psychometric review", un estudio sobre instrumentos de conocimiento en diabetes, en relación con la validación de una nueva herramienta *conocidiet diabetes* (49).

En una **primera etapa** se revisó la evidencia disponible en la literatura para la consecución del cuestionario y se llevaron a cabo técnicas de consenso por parte de expertos, de modo que cada enunciado fuera representativo del contenido semántico del indicador en su conjunto. Las declaraciones para cada indicador tenían que representar adecuadamente un aspecto específico del conocimiento dietético. Utilizando la técnica de consenso Delphi, los miembros del grupo de investigación, en varias rondas, propusieron, verificaron y modificaron el

contenido y la forma de cada declaración hasta alcanzar al menos un 90% de acuerdo (49).

Con el fin de elaborar el cuestionario en un formato comprensible para su administración, el equipo de investigación estableció el tema principal de cada indicador formulando una pregunta como encabezamiento para agrupar los conjuntos de ítems. Para cada uno de los 22 indicadores, se desarrollaron 4 ítems con un formato de respuesta verdadero-falso y un sistema de medición equivalente. Según este sistema, la falta de respuestas correctas correspondería a un valor de 1 en la escala Likert del indicador, una respuesta correcta a un valor de 2, dos respuestas correctas a un valor de 3, tres respuestas correctas a un valor de 4, y cuatro respuestas correctas a un valor de 5 (49).

En una **segunda etapa** se conformó un panel de 18 expertos. Para la inclusión de estos se tuvo en cuenta la formación y la experiencia clínica, basándose en los siguientes criterios: formación especializada en diabetes y/o experiencia clínica, docente y/o investigadora de al menos 5 años en la atención a personas con diabetes. Los expertos evaluaron cada ítem del cuestionario que había sido elaborado previamente por el grupo de investigación en la etapa 1, de acuerdo con las siguientes dimensiones: la relevancia de cada ítem para representar el constructo global evaluado y la relevancia de cada ítem para representar el indicador específico al que se hacía referencia. Para ello, se utilizaron dos escalas Likert de cuatro puntos, donde 1 significaba “nada relevante / representativo”, 2 “poco relevante / representativo”, 3 “bastante relevante / representativo” y 4 “muy relevante / representativo” (49).

Se elaboró una hoja de recolección de datos con todos los ítems generados en el proceso anteriormente mencionado y se editó en una aplicación web. El acceso a esta aplicación se facilitó mediante un hipervínculo enviado por correo electrónico a los expertos participantes. Una vez recopiladas todas las puntuaciones de los expertos, se pidió a estos que hicieran una evaluación general del cuestionario y comentaran su viabilidad y aplicabilidad práctica. El IVC de cada elemento para evaluar el constructo general se calculó dividiendo el número de evaluaciones de expertos de 3 y 4 puntos por el número total de evaluaciones realizadas para cada elemento, y asimismo se evaluó el índice de representatividad (IR) de cada ítem dentro de su respectivo indicador. Una vez

preparada la nueva versión del cuestionario, se envió de vuelta a los expertos y posteriormente se repitió el proceso anterior (49).

En la **tercera y última etapa** se elaboró una prueba piloto en la que además de contestar el cuestionario, se solicitó a los pacientes participantes en esta fase que indicaran el tiempo de cumplimentación y redactaran aquellas preguntas y frases que no hubieran comprendido. Posteriormente, se calcularon los porcentajes de finalización y los porcentajes de aciertos para cada ítem y para cada indicador (49).

2.2 VALIDEZ ESTRUCTURAL

Esta propiedad psicométrica certifica que las puntuaciones resultantes en cuanto a las respuestas del instrumento deben ser empleadas y consideradas como la medición válida del fenómeno estudiado (6,7,21,43). Mide el grado en que el instrumento manifiesta debidamente la teoría subyacente del constructo o fenómeno el cual se pretende medir y, por consiguiente, la medida concuerda con la de otros instrumentos que miden una condición similar (6,7,21,43,50,51).

Para la evaluación de la validez estructural, se realiza en primer lugar un EFA, con el propósito de revelar la estructura interna de un número relativamente amplio de variables. A través de dicho análisis se evalúa el ajuste del modelo factorial y la idoneidad de los ítems y la muestra evaluada, para lo cual se emplean 2 test; el test de esfericidad de Barlet, siempre y cuando este sea significativo ($p < 0.05$), y el test de Kaiser-Meyer-Olkin (KMO), para valores aceptables > 0.5 (52).

Posteriormente, se utilizan como valores de referencia para la extracción de dichos factores y para determinar la validez estructural del instrumento el autovalor > 1 , un valor $> 10\%$ de la varianza y el scree plot (gráfico de pantalla de factores rotados) (52,53) (figura 1). Para dicho análisis se calculan las cargas factoriales de las variables y se usa la estimación de máxima verosimilitud (ML) para evaluar el ajuste del modelo. Este método garantiza que no será descartada la posibilidad de expresar un determinado elemento en un factor o en más de un factor, todo ello a través del programa estadístico informático SPSS.

Solo aquellos ítems cuyas cargas factoriales en el EFA son ≥ 0.30 , punto de corte aceptable (54), se conservan para análisis posteriores. Aquellos ítems con valores < 0.30 deben ser eliminados, ya que no son significativos para este nuevo cuestionario, al igual que ocurre con los elementos con cargas cruzadas significativas.

Para concluir, se realiza un CFA a través del programa de ecuaciones estructurales SPSS AMOS (figura 2) para establecer si el número de factores obtenidos y sus cargas se corresponden con lo que deberíamos esperar sobre una teoría previa acerca de los datos. Los índices de ajuste del modelo incluyen chi cuadrado (χ^2), el error cuadrático medio de aproximación (RMSEA), que representa el ajuste anticipado con el valor total de la población y ya no con el de la muestra, y el índice de ajuste comparativo (CFI), que sirve para establecer si hay un buen ajuste del modelo. Para RMSEA valores < 0.06 reflejan un ajuste aceptable (55,56), mientras que para CFI valores ≥ 0.80 se consideran satisfactorios, y ≥ 0.95 reflejan un ajuste excelente (57,58).

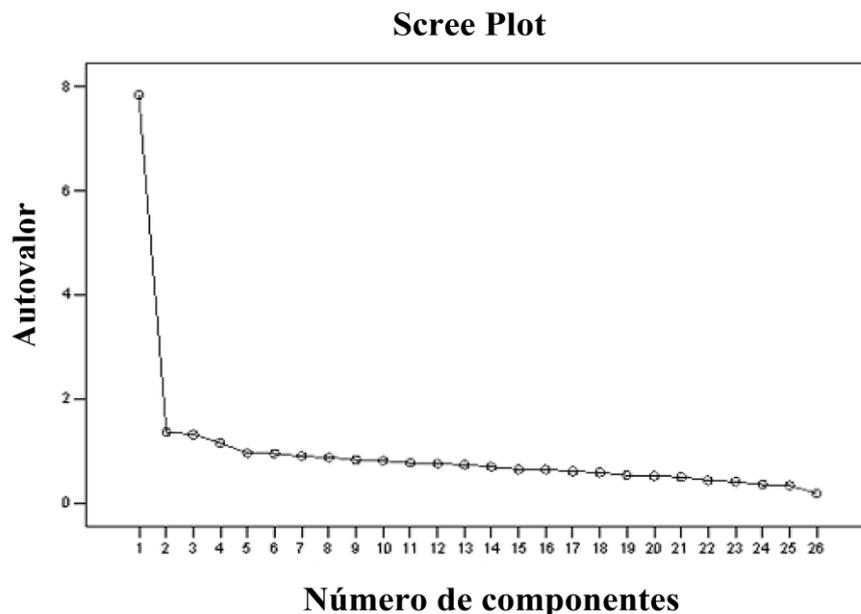


Figura 1. Scree plot (gráfico de pantalla de factores rotados).

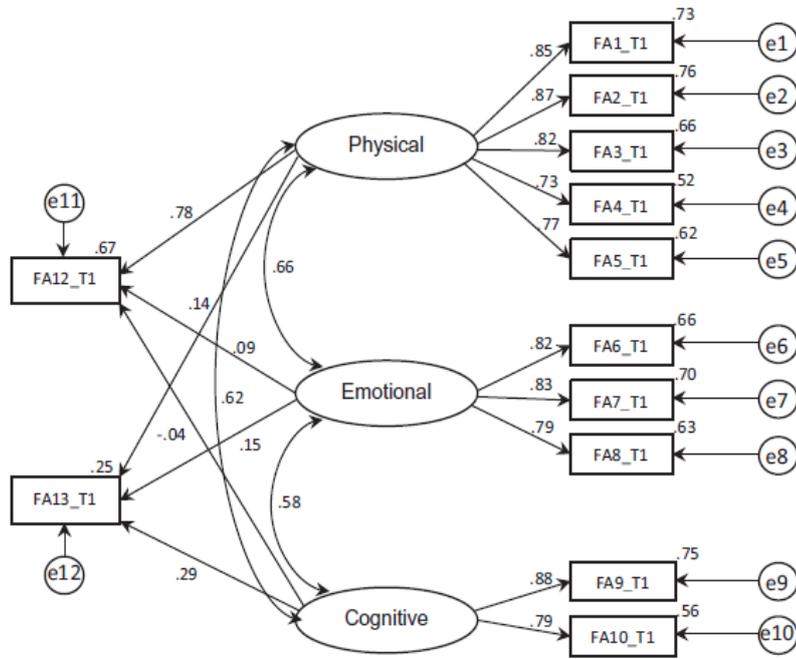


Figura 2. Análisis factorial confirmatorio (CFA) a través del programa de ecuaciones estructurales SPSS AMOS.

2.3 CONSISTENCIA INTERNA

Se refiere al grado de interrelación que existe entre los ítems de un instrumento o entre los ítems que conforman una dimensión en las escalas multidimensionales (7,43). Se valora si los ítems que evalúan una misma dimensión muestran homogeneidad entre ellos, indicativo de que los puntos de cada uno de los dominios calculan el concepto que se quiere medir y no otro diferente (59–61). Para ello es muy importante considerar que las escalas están concebidas para medir por separado los distintos dominios que forman un constructo en concreto, por lo que tenemos que calcular la consistencia interna de cada una de ellas.

Aquellas en las que la consistencia interna es alta (sus ítems miden un único constructo y este es homogéneo), aseguran en la suma de las puntuaciones de sus ítems una relación lineal con el constructo medido (61,62). A nivel estadístico la consistencia interna se puede calcular a partir del rango de los coeficientes de correlación de Pearson, del método de división por mitades de Spearman, de la fórmula de Kuder-Richardson (KR-20) para escalas dicotómicas y/o a través del α de Cronbach para escalas policotómicas (63).

Alfa de Cronbach es el método estadístico más utilizado para medir la fiabilidad de un instrumento (33,64,65). Su valor depende del número de ítems que constituyen la escala y de la correlación media entre ellos. Sus valores están comprendidos entre 0 y 1 (66-68); los valores próximos a 1 mostrarían una correlación entre ítems cercanos a la perfección, mientras que los valores de 0 no mostrarían ninguna correlación, y un coeficiente de -1 indicaría una correlación negativa entre los ítems. Es probable distinguir coeficientes negativos, supuestamente raros, cuando se lleva a cabo el cálculo en escalas que se componen de solo 2 ítems (66,69,70). De forma adicional, debe calcularse el coeficiente alfa de Cronbach cuando el instrumento está formado por un grupo de dominios (sub-escalas) para los ítems de cada uno de los dominios con respecto al valor de las puntuaciones del mismo (correlación ítem-dominio) (21). El valor mínimo admitido para alfa de Cronbach es de 0.70 (valores inferiores demuestran que la fiabilidad de la escala empleada es baja), mientras que el valor máximo deseado es de 0.90 (valores superiores revelan que hay duplicación, o lo que es lo mismo, el mismo elemento de un dominio o constructo está siendo evaluado por diferentes ítems). Los valores ideales para alfa de Cronbach estarían, por tanto, entre 0.80 y 0.90 (71).

2.4 FIABILIDAD

Describe el grado de estabilidad y consistencia en las puntuaciones conseguidas a lo largo de continuados procesos de medición utilizando un mismo instrumento (6,7,43). Desde el punto de vista estadístico, el coeficiente de correlación intraclase (ICC) es el índice más usado y adecuado para ponderar la concordancia entre distintas mediciones de una variable adquiridas a través de diferentes evaluadores o instrumentos de medida. Los valores pueden fluctuar

entre 0 y 1, donde 0 muestra una ausencia de concordancia y 1 nos indica una fiabilidad o concordancia absoluta de los resultados conseguidos. Valores de ICC < 0.4 constituyen una baja fiabilidad, valores entre 0.4 y 0.69 muestran una fiabilidad entre regular y aceptable, y valores ≥ 0.70 expresan una fiabilidad excelente (35,68).

2.4.1 Fiabilidad intra-observador o fiabilidad test-retest

Alude a la repetibilidad del instrumento y hace referencia a que, al aplicar un mismo instrumento por los mismos evaluadores empleando el mismo método, con la misma población, en dos períodos de tiempo diferentes, se consiguen puntuaciones equivalentes (6,21). Para valorar esta propiedad se puede utilizar el ICC, Pearson o Spearman, dependiendo de si la muestra es o no paramétrica (18). Al igual que el ICC, el coeficiente de Pearson se emplea para calcular la correlación entre variables cuantitativas, en caso de que las variables sean cualitativas ordinales es más conveniente la correlación de Spearman (72,73).

Se debe tener en cuenta que va a existir una variación entre lo que se está midiendo durante el tiempo transcurrido entre la primera aplicación de la escala (test) y la segunda (retest), por lo que es recomendable que el proceso entre ambas mediciones no sea demasiado largo, con el fin de evitar alteraciones en el tema o asunto de interés medido; ni tampoco muy corto, ya que podría darse lo que se conoce como *efecto de aprendizaje*, es decir, acordarse de las respuestas dadas la primera vez que se pasó el instrumento. En ambos casos, el valor de la repetibilidad se vería alterado. Otro problema que se podría presentar sería que algunos participantes no aceptasen una segunda aplicación de dicho instrumento (72-74).

2.4.2 Fiabilidad inter-observador

Hace referencia al consenso que existe entre distintos evaluadores que valoran a los mismos individuos, empleando un mismo instrumento y en un mismo periodo de tiempo. En instrumentos donde un mismo sujeto es el que aporta las respuestas (cuestionarios autoadministrados) esta propiedad no es medible, ya que no se produce la intervención de los evaluadores en los

resultados del mismo. Existen algunas limitaciones en lo que se refiere a esta medición: una sería la presencia de acuerdo por azar entre los evaluadores y, otra, la aparición de un error sistemático en alguno de ellos (21,74). En el caso de que sea precisa su evaluación, los métodos estadísticos empleados son idénticos a los utilizados para la fiabilidad test-retest.

2.5 ERROR DE MEDICIÓN

Un error de medición es la diferencia entre la puntuación real u obtenida y la puntuación real teórica (6). El error estándar de medición (SEM) es uno de los métodos estadísticos más utilizados para medir el error de medición de un instrumento. Se trata de una determinación de la cantidad de variación o dispersión en los errores de medición para una prueba, y constituye un gran número de métodos estadísticos, como el análisis de la varianza (ANOVA), el análisis factorial o el análisis de regresión múltiple. También se emplea para responder a cuestiones que requieren métodos de análisis de regresión múltiple; por ejemplo, a través del SEM o ANOVA podemos analizar comparaciones de dos grupos o más, o comparaciones intra-grupos llegando indistintamente al mismo resultado (7,43,75)

2.5.1 Parámetros estimados del SEM

Los distintos parámetros deben ser especificados y estimados, tanto para un modelo de medición como para un modelo estructural. Una vez especificados, estos pueden ser libres, fijos o restringidos. Un parámetro libre es aquel que se calcula a partir del análisis y tiene un valor desconocido. Un parámetro fijo es aquel que es pre-especificado por el investigador para adquirir un valor determinado (en la mayoría de los casos suele ser 10). Finalmente, un parámetro restringido es aquel que posee también un valor desconocido, pero el investigador le asigna un valor determinado (76,77).

2.5.2 Etapas del SEM

Está constituido por 4 etapas: una primera etapa de especificación, una segunda etapa de estimación, una tercera etapa de evaluación de ajuste global y una última etapa de reespecificación.

- *Especificación del modelo:* durante esta primera etapa se constituyen las hipótesis sobre la relación entre las distintas variables. Se debe deliberar si las relaciones son unidireccionales o bidireccionales, fundamentándose en la evidencia previa, en la teoría o en ambas (75). Dichas relaciones pueden ser directas o indirectas, siempre que en el modelo se incorporen variables que intercedan en el efecto de una variable sobre la otra (78,79). Se trata de una técnica más de tipo confirmatorio que exploratorio.

- *Estimación del modelo:* una vez que el modelo haya sido especificado e identificado (un modelo se considera identificado siempre que el número de parámetros estimados no supere al número de datos observados, y solo los modelos que se encuentran identificados pueden ser estimados) sus parámetros son evaluados con la finalidad de reducir al máximo las diferencias entre la matriz de covarianza observada y la matriz de covarianza poblacional (80).

Aunque existen multitud de procedimientos de estimación aptos para su uso, el ML y los métodos de mínimos cuadrados (SL) (81) y de distribución asintótica libre (ADF) (82) son los más conocidos.

- *Evaluación del ajuste global del modelo:* de forma sistemática, el ajuste del modelo es calculado mediante el chi cuadrado (χ^2), bastante sensible a tamaños muestrales grandes. Por ello, la evaluación global del modelo habitualmente se lleva a cabo teniendo en cuenta otros indicadores de bondad de ajuste (tienen un rango que oscila entre 0 y 1), con valores altos indicando una mayor varianza expuesta por el modelo. El CFI y el índice de Tucker- Lewis (TLI) son habitualmente muy utilizados y contrastan el modelo existente con uno nulo. Los valores de CFI > 0.95 y TLI > 0.90, son indicativos de un buen ajuste (58).

Un buen ajuste de los datos al modelo es obtenido también a partir de valores de residuos próximos a 0, lo cual revelaría el porcentaje de varianza no explicado por

el modelo. De estos, son habitualmente reportados el RMSEA y la raíz cuadrada media residual estandarizada (SRMR), con valores para RMSEA < 0.6 y SRMR < 0.08 cuya consideración es aceptable (56).

- *Re-especificación del modelo*: una vez valorado el ajuste global del modelo a los datos, podemos optar por dos opciones: la primera consistiría en la aclaración e información sobre los resultados y la segunda sobre la re-especificación del modelo. Tenemos que tener en cuenta que la primera opción es siempre la recomendable, pero la estimación del ajuste global del modelo a los datos no es siempre la esperada, precisándose una re-especificación del modelo. Algunas de las razones para esta re-especificación encontrarían su justificación en la búsqueda de algún patrón de correlación inicialmente excluido, el rastreo de algún residuo muy valioso o el manejo de algún algoritmo estadístico (75).

2.6 VALIDEZ DE CRITERIO CONVERGENTE

A través de esta propiedad se establecen las correlaciones de las puntuaciones obtenidas mediante escalas diferentes (7,43). Cuando los resultados entre ambas medidas presentan correlaciones significativas al ser comparados con otros instrumentos que cuantifican el mismo constructo, se dice que *convergen*, aceptándose que ambas escalas son conceptualmente análogas o congruentes. Por el contrario, si se obtienen correlaciones bajas o negativas al comparar las puntuaciones de escalas que evalúan constructos diferentes, indica que las escalas *divergen*, mostrando una asociación no significativa entre ambas variables, lo que viene a confirmar que evalúan constructos diferentes. En caso contrario, sería indicativo de que la escala que se pretende validar no es lo bastante específica para medir el constructo de interés en la población seleccionada (18,59,83,84). Dicha propiedad habitualmente se calcula empleando el coeficiente de correlación de Pearson, esperando conseguir en circunstancias idóneas unos resultados psicométricos adecuados, que suponen una correlación > 0.60 (convergente) o una correlación < 0.20 (divergente) (85).

2.7 VALIDEZ DE CRITERIO

Determina el grado en el que las puntuaciones obtenidas a partir de una escala son válidas al ser comparadas con un estándar o un patrón de referencia (6,7,21,43). Así pues, el nuevo instrumento que se pretende medir debe compararse con una escala ya existente, considerablemente reconocida y que haya justificado ser el instrumento más adecuado para medir dicho fenómeno de interés (“gold standard”). De tal modo se comparan las puntuaciones alcanzadas con cada una de las diferentes escalas para valorar si existe una correlación adecuada entre ambas (21,86).

Siempre que se disponga de un estándar o de una escala alternativa que efectúe la misma función a tales efectos (se muestre fiable, independiente, válida y sea capaz de medir la misma condición de interés), conviene proceder a ejecutar los siguientes mismos pasos para valorar dicha propiedad: elegir el estándar o su equivalente más adecuado, seleccionar a partir de la población objeto de estudio una muestra representativa, emplear la escala en evaluación y conseguir un puntaje para cada individuo y, finalmente, evaluar a cada uno de los sujetos con el estándar, para posteriormente comparar las puntuaciones obtenidas con ambos instrumentos (21,74).

Pueden evaluarse las características de esta propiedad, y dependiendo del momento en que se realice la comparación de resultados, la validez será concurrente o predictiva (86,87). La validez concurrente determina el grado de correlación existente entre los resultados alcanzados por la escala considerada “criterio” o “estándar” y la escala que está siendo evaluada cuando ambas son administradas simultáneamente (86). Dicha comparación se lleva a cabo estadísticamente utilizando el coeficiente de correlación de Pearson (21,88) o de Spearman en función de las características y como se encuentren distribuidos los datos (87). Por otro lado, la validez predictiva valora el grado en que la nueva escala de medición es competente para predecir las puntuaciones obtenidas por el “gold standard” cuando este no es aplicado en el mismo espacio de tiempo, sino en algún punto diferente del futuro (86). Desde un punto de vista estadístico, esta comparación se efectúa de igual manera que en la validez concurrente.

La finalidad de la validez de criterio es que se dé una correlación adecuada entre ambos instrumentos. Siempre debe tenerse en cuenta que si el fin para la

validación de una escala nueva está fundamentado en que esta presente mayor utilidad (sencillez en su aplicación, proceso de calificación, economía, comodidad para el individuo, pertinencia y disminución del error de medida), la validación de criterio precisará la consecución de correlaciones ≥ 0.8 , indicativas de que las dos escalas son psicométricamente similares, mientras que si el propósito es demostrar que la nueva versión es aún más válida y mejor que el instrumento de referencia, lo correcto sería conseguir correlaciones en un rango entre 0.3 y 0.7, lo que revelaría que ambos instrumentos son distintos aunque midan el mismo atributo (87–89).

2.8 PRUEBA DE HIPÓTESIS

Cuando no existe un “gold standard” para evaluar el instrumento, la forma comúnmente utilizada para investigar la validez de la nueva herramienta es probar hipótesis sobre relaciones esperadas con otras medidas de resultados de buena calidad y/o diferencias esperadas entre grupos relevantes. Es de gran importancia definir hipótesis de antemano al evaluar la validez de constructo de la nueva herramienta para que los autores puedan sacar conclusiones no sesgadas después de la recopilación y el análisis de los datos (7).

La prueba de hipótesis es un proceso iterativo continuo (90). Las hipótesis específicas deben incluir una indicación de la dirección esperada y la magnitud de las correlaciones o diferencias, por lo que la prueba de hipótesis determinará si la dirección y la magnitud de una correlación o diferencia es similar a lo que podría esperarse en función de los constructos que se están midiendo. Cuantas más hipótesis se prueben sobre si los datos corresponden a hipótesis formuladas *a priori*, más evidencia se recopilará para la validez de constructo.

2.9 SENSIBILIDAD AL CAMBIO

La sensibilidad al cambio es la propiedad psicométrica presente en un instrumento para descubrir los cambios a través del tiempo en la realidad en la que se pretende medir, todo ello como respuesta de un mismo individuo o varios individuos sobre dicho constructo (7,43). Dicha propiedad está frecuentemente presente en escalas diagnósticas, mediciones prospectivas o ensayos clínicos, en

los que la sensibilidad al cambio y la especificidad posibilitan valorar la respuesta a una intervención o un tratamiento. No se suele incluir en estudios que contienen variables como la *satisfacción*, el *bienestar*, las *actitudes* o las *percepciones* (6,21,62,91,92).

2.10 VALIDEZ TRANSCULTURAL

Se refiere al grado en que la interpretación de los ítems en un instrumento traducido o adaptado culturalmente es un reflejo adecuado del desempeño de los ítems de la versión original del instrumento (7,43).

El proceso de validación transcultural es preciso efectuarlo siempre que el instrumento en cuestión sea adaptado a un tipo de población específico, aunque la escala ya se encuentre traducida en su idioma original en cuestión. Es preciso llevar a cabo un proceso de traducción que posteriormente tiene que ser revisado por un comité de expertos (7,43,93), pasar una prueba piloto (pre-test) y, en última instancia, someterse a un proceso de validación.

Dicho proceso de traducción debe ser realizado por traductores expertos, preferiblemente 2 traductores nativos, uno de ellos experto en el tema. Ambos trabajarán de forma independiente y a través de un enfoque de traducción secuencial hacia delante y hacia atrás (para llevar a cabo el proceso de traducción hacia atrás, al igual que ocurriera en el proceso de traducción hacia delante, es necesario que ambos traductores sean nativos y que trabajen de forma independiente, pero en este caso lo ideal es que no tengan conocimientos sobre el tema propuesto) y finalmente, redacten un informe con los principales comentarios sobre el proceso de traducción (7,43).

Posteriormente, se constituye un panel de expertos para evaluar la validez de contenido del instrumento a través de una serie de indicadores habilitados para este tipo de procesos, con el fin de asegurar que el nuevo instrumento es considerado un reflejo semántico correcto de la versión en inglés y abarca todo su contenido original. Una vez obtenida la versión modificada del nuevo instrumento, el siguiente paso es someter dicha herramienta a una prueba piloto (pre-test) en una población de pacientes que represente a la población objetivo, con el fin de conocer la calidad metodológica del nuevo instrumento en cuanto a relevancia, exhaustividad y comprensibilidad con respecto al constructo de

interés y a la población de estudio, y si definitivamente el contenido adquirido en dicho instrumento es un reflejo adecuado del constructo que se pretende medir (7,43).

Una vez obtenida la versión de la escala traducida y adaptada culturalmente, se somete a un proceso de validación a través de la medición de una serie de propiedades psicométricas (3,5-7,43,48,94).

Normativa COSMIN

3. NORMATIVA COSMIN

3.1 LISTA DE VERIFICACIÓN COSMIN

La lista de verificación de los Estándares basados en el COnsenso para la selección de INstrumentos de Medición de la salud (COSMIN) es una herramienta desarrollada para evaluar la calidad metodológica de estudios individuales sobre las propiedades de medición de los cuestionarios de evaluación, y tiene como objetivo facilitar la selección de un instrumento de alta calidad para la investigación y la práctica clínica (3,6).

Dentro de esa lista de verificación COSMIN existen dos conceptos fundamentales: PRO y PROM. Un resultado informado por el paciente (PRO) es cualquier aspecto del estado de salud de un paciente que es evaluado directamente por él mismo sin que nadie que no sea él interprete su propio resultado (1). Los PRO se evalúan con mayor frecuencia mediante cuestionarios autoadministrados, también conocidos como medidas de resultado informadas por el paciente (PROM). Sin embargo, es sabido que la calidad de estos cuestionarios varía considerablemente y, por lo general, no es fehaciente si se ha seleccionado la herramienta más fiable y válida (95).

Otro aspecto esencial es que COSMIN incluye un cuadro para cada propiedad de medición que contiene unos estándares que hacen referencia a requisitos en cuanto al diseño y selección de métodos estadísticos, con el fin de evaluar la calidad del estudio (43,94).

Cuando se emplea COSMIN en las revisiones sistemáticas, estas se convierten en herramientas importantes para seleccionar el cuestionario más adecuado con la pretensión de medir un constructo de interés en una población de estudio específica. Esto es importante, ya que el número de estas revisiones ha aumentado de apenas una por año a principios de la década de los 90 a más de 100 cada año en la actualidad (96), produciéndose un deterioro en cuanto a calidad se refiere, algo que queda de manifiesto en una revisión reciente donde se evidenció que existe un margen considerable de mejora (97).

Además, estas revisiones, donde se hace uso de la lista de verificación COSMIN, pueden proporcionar una descripción general completa de las propiedades de medición de dichos cuestionarios, respaldar las recomendaciones basadas en la evidencia para la selección de la herramienta más adecuada para un propósito determinado (investigación o práctica clínica, aplicaciones discriminatorias, evaluativas o predictivas) e identificar lagunas en el conocimiento sobre las propiedades psicométricas relacionadas con estos instrumentos de evaluación, que pueden utilizarse para diseñar nuevos estudios sobre las propiedades de medición (5).

En ausencia de una guía extensa y publicada para revisiones sistemáticas sobre cuestionarios de evaluación, el comité directivo de COSMIN desarrolló un protocolo para revisiones sistemáticas sobre este tipo de cuestionarios, que está disponible en la página web de COSMIN desde 2011 (<http://www.cosmin.nl>) (4).

En unos 10 puntos, la guía COSMIN describe la metodología de revisiones sistemáticas sobre cuestionarios existentes, para los cuales se dispone de al menos alguna información sobre sus propiedades de medición, su uso con fines evaluativos y su selección para un propósito específico. Se puede encontrar información detallada que respalda la realización de una revisión sistemática en la guía “COSMIN methodology for systematic reviews of PROMs—user manual”, así como en “COSMIN methodology for assessing the content validity of PROMs—user manual”, disponible en el sitio web de COSMIN (4,7,8).

Este manual se desarrolló de acuerdo con las guías existentes para revisiones: el manual Cochrane para revisiones sistemáticas de intervenciones (98) y para revisiones de precisión de pruebas diagnósticas (99), la Declaración PRISMA (100), la lista de verificación STARD (101), las normas del Instituto de Medicina para revisiones sistemáticas de la investigación de la eficacia comparativa (102), los criterios actualizados para las buenas propiedades de medición de Terwee y Prinsen (7,8) y la metodología para formular y calificar recomendaciones en la práctica clínica GRADE (7,8,103).

3.2 GUÍA COSMIN PARA REVISIONES SISTEMÁTICAS

Para la realización de una revisión sistemática con el objetivo de medir un constructo de interés en una población de estudio específica, se recomienda seguir un procedimiento basado en 10 puntos. Estos a su vez se subdividen en 3 apartados: A, B y C.

(A. Búsqueda de la literatura.

Este apartado consta de 4 puntos (del 1 al 4). Se trata de procedimientos estándar que cumplen con las directrices existentes cuando se realizan revisiones sistemáticas (98,99).

3.2.1 Punto 1. Formular el objetivo de la revisión.

El objetivo de una revisión sistemática para medir un constructo de interés en una población de estudio específica en relación con cuestionarios de evaluación se centra en la calidad de estos. Debe incluir los siguientes cuatro elementos clave: el constructo o qué medir, la población, el tipo de instrumento y las propiedades de medición de interés (5).

3.2.2 Punto 2. Establecer los criterios de elegibilidad.

Los criterios de elegibilidad deben estar de acuerdo con los cuatro elementos clave del objetivo de la revisión: el cuestionario, que debe medir el constructo de interés; la muestra del estudio, que debe representar la población de interés; el estudio, que debe hacer referencia al cuestionario, y el objetivo del estudio, que debe ser la evaluación de una o más propiedades de medición, el desarrollo de un cuestionario (para calificar la validez de contenido) o la evaluación de la interpretabilidad de esos cuestionarios de interés (5).

Se recomienda la exclusión de los estudios que solo utilizan un cuestionario como instrumento de medición de resultados (aquellos en los que se usa el cuestionario para medir el resultado (ensayos controlados aleatorios) o como validación de otro instrumento) porque proporcionan evidencia indirecta sobre las propiedades de medición del mismo. Finalmente, se aconseja incluir

únicamente artículos a texto completo porque, a menudo, en los resúmenes se encuentra información muy limitada sobre el diseño de un estudio, lo que obstaculizará la evaluación de la calidad del mismo y los resultados de las propiedades de medición en los puntos posteriores (puntos del 5 al 7) (5).

3.2.3 Punto 3. Realizar una búsqueda bibliográfica.

De acuerdo con la metodología Cochrane (98,99) y por consenso, MEDLINE y EMBASE son las bases de datos mínimas en las que se deben efectuar búsquedas. Además, se recomienda buscar en otras bases de datos específicas de contenido según el constructo y la población de interés (104).

Una estrategia de búsqueda adecuada consiste en una colección completa de términos de búsqueda (es decir, términos de índice y palabras de texto libre) para los cuatro elementos clave del objetivo de la revisión: constructo, población, tipo de instrumento y propiedades de medición (104) (figura 3).

Con respecto a los términos de búsqueda para las propiedades de medición, se recomienda utilizar un filtro de búsqueda validado y altamente sensible, disponible para PubMed y EMBASE, en el sitio web de COSMIN (4).

De acuerdo con la metodología Cochrane se recomienda buscar en las bases de datos desde la fecha de inicio hasta la actualidad. El caso del uso de restricciones de idioma va a depender de los criterios de inclusión definidos en el punto 2. Aun así, se aconseja no emplearlos en la estrategia de búsqueda incluso si no hay recursos para traducir los artículos, de esta manera los revisores pueden al menos informar de su existencia (98,99).

Search strategy for Pubmed

- 1 "nutritional knowledge" [All Fields] OR nutrition [All Fields] OR "nutrition status" [All Fields] OR "dietary knowledge" [All Fields] OR diet [All Fields]
"nutritional status"[All Fields] OR "nutrition"[All Fields] OR "nutrition status"[All Fields] AND "questionnaire"[All Fields] AND "assessment tool"[All Fields] AND "diabetes"[All Fields] AND "humans"[MeSH Terms]
- 2 "Diabetes mellitus" [MeSH] OR diabetes [All Fields]
- 3 (Instrument[tiab] OR instruments[tiab] OR measure [tiab] OR measures [tiab] OR questionnaire[tiab] OR questionnaires[tiab] OR scale[tiab] OR scales[tiab] OR tool[tiab] OR tools[tiab] OR survey [tiab] OR test [tiab])
- 4 Instrumentation[sh] OR methods[sh] OR "Validation Studies"[pt] OR "Comparative Study"[pt] OR "psychometrics"[MeSH] OR psychometr*[tiab] OR clinimetr*[tw] OR clinometr*[tw] OR "outcome assessment (health care)"[MeSH] OR "outcome assessment"[tiab] OR "outcome measure*" [tw] OR "observer variation"[MeSH] OR "observer variation"[tiab] OR "Health Status Indicators"[Mesh] OR "reproducibility of results"[MeSH] OR reproducib*[tiab] OR "discriminant analysis"[MeSH] OR reliab*[tiab] OR unreliab*[tiab] OR valid*[tiab] OR coefficient[tiab] OR homogeneity[tiab] OR homogeneous[tiab] OR "internal consistency"[tiab] OR (cronbach*[tiab] AND (alpha[tiab] OR alphas[tiab])) OR (item[tiab] AND (correlation*[tiab] OR selection*[tiab] OR reduction*[tiab])) OR agreement[tiab] OR precision[tiab] OR imprecision[tiab] OR "precise values"[tiab] OR test-retest[tiab] OR (test[tiab] AND retest[tiab]) OR (reliab* [tiab] AND (test[tiab] OR retest[tiab])) OR interrater[tiab] OR inter-rater[tiab] OR intrarater[tiab] OR intra-rater[tiab] OR intertester[tiab] OR inter-tester[tiab] OR intratester[tiab] OR intra-tester[tiab] OR interobserver[tiab] OR inter-observer[tiab] OR intraobserver[tiab] OR intraobserver[tiab] OR intertechnician[tiab] OR inter-technician[tiab] OR intratechnician[tiab] OR intra-technician[tiab] OR interexaminer[tiab] OR inter-examiner[tiab] OR intraexaminer[tiab] OR intra-examiner[tiab] OR interassay[tiab] OR inter-assay[tiab] OR intraassay[tiab] OR intra-assay[tiab] OR interindividual[tiab] OR inter-individual[tiab] OR intraindividual[tiab] OR intra-individual[tiab] OR interparticipant[tiab] OR inter-participant[tiab] OR intraparticipant[tiab] OR intra-participant[tiab] OR kappa[tiab] OR kappa's[tiab] OR kappas[tiab] OR repeatab*[tiab] OR ((replicab*[tiab] OR repeated[tiab]) AND (measure[tiab] OR measures[tiab] OR findings[tiab] OR result[tiab] OR results[tiab] OR test[tiab] OR tests[tiab])) OR generaliza*[tiab] OR generalisa*[tiab] OR concordance[tiab] OR (intraclass[tiab] AND correlation*[tiab]) OR discriminative[tiab] OR "known group"[tiab] OR "factor analysis"[tiab] OR "factor analyses"[tiab] OR dimension*[tiab] OR subscale*[tiab] OR (multitrait[tiab] AND scaling[tiab] AND (analysis[tiab] OR analyses[tiab])) OR "item discriminant"[tiab] OR "interscale correlation*" [tiab] OR error[tiab] OR errors[tiab] OR "individual variability"[tiab] OR (variability[tiab] AND (analysis[tiab] OR values[tiab])) OR (uncertainty[tiab] AND (measurement[tiab] OR measuring[tiab])) OR "standard error of measurement"[tiab] OR sensitiv*[tiab] OR responsive*[tiab] OR ((minimal[tiab] OR minimally[tiab] OR clinical[tiab] OR clinically[tiab]) AND (important[tiab] OR significant[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR (small*[tiab] AND (real[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR "meaningful change" [tiab] OR "ceiling effect"[tiab] OR "floor effect"[tiab] OR "Item response model"[tiab] OR IRT[tiab] OR Rasch[tiab] OR "Differential item functioning"[tiab] OR DIF[tiab] OR "computer adaptive testing"[tiab] OR "item bank"[tiab] OR "cross-cultural equivalence"[tiab]
- 5 #1 AND #2 AND #3 AND #4
- 6 "Protocol"[Publication Type] OR "addresses"[Publication Type] OR "biography"[Publication Type] OR "case reports"[Publication Type] OR "comment"[Publication Type] OR "editorial"[Publication Type] OR "congresses" [Publication Type] OR "consensus development conference"[Publication Type] OR "consensus development conference"[Publication Type] OR "practice guideline"[Publication Type] OR "suffering from"[tiab] OR "animals"[MeSH]
- 7 #5 NOT #6
- 8 FILTER: Language (English and Spanish)
- 9 FILTER: Species (Humans)

Figura 3. Estrategia de búsqueda en Pubmed.

3.2.4 Punto 4. Seleccionar resúmenes y artículos a texto completo.

Se recomienda que la selección se lleve a cabo por 2 revisores de forma independiente; si un estudio parece relevante por al menos un revisor basándose en el resumen (o en caso de duda), el artículo a texto completo debe recuperarse y revisarse. Deben discutirse las diferencias y, si no se llega a un consenso entre los dos revisores, se recomienda consultar a un tercer revisor. Se aconseja verificar todas las referencias de los artículos incluidos para buscar estudios adicionales potencialmente notorios. Si se encuentran muchos artículos nuevos es probable que la estrategia de búsqueda inicial no haya sido lo suficientemente completa, por lo que posiblemente deba mejorarse o repetirse (98,99) (figura 4).

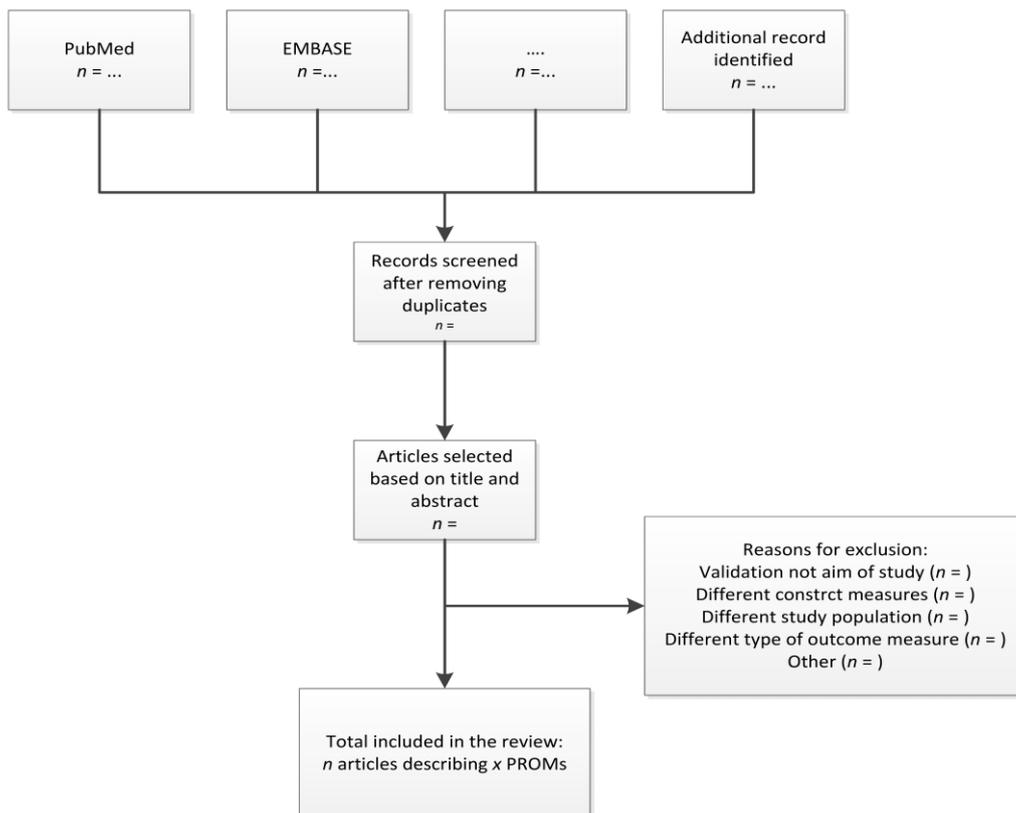


Figura 4. Diagrama de flujo COSMIN del proceso de búsqueda.

(B. Evaluación de las propiedades de medición)

Este apartado consta de 3 puntos (del 5 al 7) y hace referencia a la evaluación de las propiedades de medición de los cuestionarios incluidos. A su vez consta de 3 pasos:

En primer lugar, la calidad metodológica de cada estudio sobre una propiedad de medición se evalúa utilizando la lista de verificación COSMIN Risk of Bias y cada estudio se califica como de “muy buena”, “adecuada”, “dudosa” o “inadecuada” calidad (94).

En segundo lugar, la calidad de los resultados para cada estudio sobre una propiedad de medición se califica en función de los criterios actualizados para las buenas propiedades de medición de Terwee y Prinsen, pudiendo ser evaluado como “suficiente” (+), “insuficiente” (-) o “indeterminado” (?) para la totalidad de las propiedades psicométricas, excepto la validez de contenido que se puntúa de forma diferente, como “suficiente” (+), “insuficiente” (-) o “inconsistente” (\pm) (7,8).

En tercer lugar, la evidencia se resumirá y la calidad de esta se calificará mediante el método GRADE (7,8,103).

Los resultados de todos los estudios disponibles sobre una propiedad de medición se agrupan cuantitativamente o se resumen cualitativamente y se comparan con los criterios de buenas propiedades de medición, para determinar si la propiedad de medición del cuestionario es “suficiente” (+), “insuficiente” (-), inconsistente (\pm), o indeterminada (?) (7,8).

Esas calificaciones generales para cada propiedad de medición irán acompañadas de una calificación de la calidad de la evidencia. El método GRADE para revisiones sistemáticas de estudios de intervención específica ostenta cuatro niveles para calificar esa calidad de la evidencia (evidencia de calidad “alta”, “moderada”, “baja” o “muy baja”), dependiendo de la presencia de 5 factores: riesgo de sesgo, ausencia de evidencia directa, inconsistencia, imprecisión y sesgo de publicación (7,8,103).

Para las revisiones sistemáticas se emplea un método GRADE modificado que se utiliza para degradar la calidad de la evidencia cuando existen preocupaciones sobre la fiabilidad de los resultados. Para ello se han adoptado 4 de esos 5 factores GRADE para su uso en la metodología COSMIN: riesgo de

sesgo (calidad metodológica de los estudios), inconsistencia (incongruencia inexplicable de los resultados entre los estudios), imprecisión (tamaño total de la muestra de los estudios disponibles) y ausencia de evidencia directa (evidencia de poblaciones diferentes a la población de interés en la revisión) (7,8,103) (figura 5).

Quality of evidence	Lower if
High	Risk of bias
Moderate	-1 Serious
Low	-2 Very serious
Very low	-3 Extremely serious
	Inconsistency
	-1 Serious
	-2 Very serious
	Imprecision
	-1 total n=50-100
	-2 total n<50
	Indirectness
	-1 Serious
	-2 Very serious

Figura 5. Método GRADE modificado para estudios de validación.

3.2.5 Punto 5. Evaluación de la validez de contenido.

La validez de contenido se refiere al grado en que el contenido de un cuestionario es un reflejo adecuado del constructo que se va a medir (3). Se considera la propiedad de medición más importante porque debe quedar claro

que los ítems del cuestionario son relevantes, exhaustivos y comprensibles con respecto al constructo de interés y la población de estudio (8).

Si hay evidencia de alta calidad de que la validez de contenido de un cuestionario es insuficiente, no se tendrá en cuenta en los puntos 6, 7 y 8 de la revisión sistemática, y se procederá a formular directamente una recomendación para esta en el punto 9 (5).

3.2.6 Punto 6. Evaluación de la estructura interna: validez estructural, consistencia interna y validez transcultural / invariancia de medición.

La estructura interna se refiere a cómo se relacionan los diferentes elementos de un cuestionario, lo cual es relevante saber para decidir cómo se pueden combinar los elementos en una escala o subescala (3). En este punto se realiza una evaluación de la validez estructural (incluida la unidimensionalidad), la consistencia interna y la validez transcultural / invariancia de medición. Estas 3 propiedades se centran en la calidad de los elementos individuales y las relaciones entre estos (5).

Es recomendable que todas las propiedades psicométricas que conforman la estructura interna sean evaluadas directamente después de la validación del contenido. Como la evidencia de la validez estructural (o unidimensionalidad) de una escala o subescala es un requisito previo para la interpretación de los análisis de consistencia interna (alfa de Cronbach), es aconsejable evaluar primero la validez estructural, seguida de la consistencia interna y, por último, la validez transcultural / invariancia de medición (5).

El punto 6 solo es relevante para los cuestionarios que se basan en un modelo reflexivo que asume que todos los elementos de una escala o subescala son manifestaciones de un constructo subyacente y se espera que estén correlacionados. Si se supone que los elementos de una escala o subescala no están correlacionados (modelo formativo), estos análisis no son fundamentales y se puede omitir este punto. Si no se informa si un cuestionario se basa en un modelo reflexivo o formativo, los revisores deben decidir sobre el contenido de este para saber si se fundamenta en un modelo u otro (105).

3.2.7 Punto 7. Evaluación del resto de propiedades.

Las propiedades de medición restantes (fiabilidad, error de medición, validez de criterio, prueba de hipótesis (validez de constructo) y sensibilidad al cambio) deben evaluarse nuevamente, siguiendo los 3 pasos descritos anteriormente en el apartado B (evaluación de las propiedades de medición).

A diferencia de la validez de contenido y la estructura interna, la evaluación de estas propiedades de medición proporciona información sobre la calidad de la escala o subescala en su conjunto, en lugar de a nivel de ítem (5).

(C. Selección del cuestionario)

Este apartado consta de 3 puntos (del 8 al 10) y se refiere a la evaluación de la interpretabilidad y viabilidad de los cuestionarios, formulando recomendaciones y proporcionando información sobre la revisión sistemática.

3.2.8 Punto 8. Descripción de la interpretabilidad y la viabilidad.

La interpretabilidad se define como el grado en que se puede asignar un significado cualitativo (connotaciones clínicas o comúnmente entendidas) a las puntuaciones cuantitativas de un cuestionario o simplemente al cambio de las puntuaciones (3); mientras que la viabilidad se detalla como la facilidad de aplicación del cuestionario en su entorno previsto, dadas las limitaciones de tiempo o económicas, y hace referencia a aspectos como el coste de un instrumento, la duración del mismo, el tipo y facilidad para ser administrado (104). Se aplica a los pacientes que completan el cuestionario (autoadministrado) y a los investigadores o médicos que entrevistan o entregan ese cuestionario a los pacientes.

Interpretabilidad y viabilidad no son propiedades de medición porque no hacen referencia a la calidad de un cuestionario. Sin embargo, se consideran elementos imprescindibles para su selección. En caso de que existan dos cuestionarios muy difíciles de diferenciar en términos de calidad, se recomienda que se tengan en cuenta los aspectos de viabilidad en la selección del instrumento

más adecuado, de tal modo que los revisores deban decidir si es factible en cuanto a tiempo (duración del mismo) y a coste (104).

3.2.9 Punto 9. Formulación de recomendaciones.

Las recomendaciones sobre el cuestionario más adecuado para su uso en una aplicación evaluativa, se formulan con respecto al constructo de interés y a la población de estudio. Para llegar a una recomendación basada en la evidencia y totalmente transparente, recomendamos clasificar los cuestionarios incluidos en tres categorías (106):

a) Cuestionarios que tienen potencial para ser recomendados como el más adecuado para el constructo y la población de interés (cuestionario con evidencia de suficiente validez de contenido (cualquier nivel) y al menos poca evidencia de suficiente consistencia interna).

b) Cuestionarios que pueden tener el potencial de ser recomendados, pero se necesitan más estudios de validación (cuestionarios categorizados no en a o c).

c) Cuestionarios que no deben recomendarse (cuestionarios con evidencia de alta calidad para una propiedad de medición insuficiente).

Se deben justificar las razones por las que un cuestionario se coloca en una determinada categoría y se tienen que dar instrucciones sobre el trabajo de validación futuro, si corresponde. Se aconseja asesorar sobre el cuestionario más adecuado. Esta recomendación no solo debe basarse en la evaluación de las propiedades de medición, sino que también puede depender de aspectos de interpretabilidad y factibilidad (104).

3.2.10 Punto 10. Información sobre la revisión sistemática.

De acuerdo con la Declaración PRISMA (100), se recomienda reportar la siguiente información:

1) Resultados de la búsqueda bibliográfica y selección de los estudios e instrumentos presentados en el diagrama de flujo PRISMA (incluyendo el número final de artículos y el número de instrumentos incluidos en la revisión).

2) Características de los instrumentos incluidos, como el nombre de estos, los constructos que se miden, la población de estudio para la que se desarrolló el instrumento, el contexto o los contextos de uso previstos, la versión lingüística del cuestionario, el número de escalas o subescalas, el número de ítems, opciones de respuesta, período de recuperación y aspectos de interpretabilidad y viabilidad.

3) Características de las poblaciones de estudio: ubicación geográfica, idioma, tipo de patología, población objetivo, tamaño de la muestra, edad, sexo, entorno y país.

4) Calidad metodológica de cada estudio por propiedad de medición e instrumento.

5) Tabla resumen con los hallazgos por propiedad de medición, incluidos los resultados agrupados o resumidos, su calificación general (suficiente (+), insuficiente (-), inconsistente (\pm) o indeterminada (?)) y la calificación de la calidad de la evidencia (alta, moderada, baja o muy baja).

Estas tablas (una por cada propiedad de medición) se utilizarán, en última instancia, para proporcionar recomendaciones y para la selección del instrumento más adecuado para un propósito determinado o un contexto de uso particular.

Criterios de Terwee y Prinsen

4. CRITERIOS DE TERWEE Y PRINSEN

Los criterios de Terwee y Prinsen se establecieron con el principal objetivo de valorar la calidad de los resultados para cada estudio sobre una propiedad de medición, en los siguientes dominios: validez de contenido, validez estructural, consistencia interna, fiabilidad, error de medición, prueba de hipótesis (validez de constructo), validez transcultural, validez de criterio y sensibilidad al cambio (figura 6). También se crearon para detectar deficiencias y lagunas en el conocimiento de estas propiedades psicométricas y para diseñar estudios de validación (5,7,104).

En un primer momento, las propiedades a evaluar diferirían de las actuales, estas eran: validez de contenido, consistencia interna, validez de criterio, validez de constructo, reproducibilidad (concordancia y fiabilidad), sensibilidad, efecto piso techo e interpretabilidad. Contenían 4 ítems que valoraban si estas guardaban relación con determinados estándares de calidad metodológica, en función de 4 criterios: suficiente (+), indeterminado (?), insuficiente (-) y no se ha encontrado información (0) (107).

En la actualidad, esos 4 criterios quedaron reducidos a 3: suficiente (+), indeterminado (?) e insuficiente (-) para la totalidad de las propiedades psicométricas (5,7,104), exceptuando validez de contenido (evaluada a partir de cuatro criterios: suficiente (+), inconsistente (\pm), insuficiente (-) e indeterminado (?)) (8).

Measurement property	Rating ¹	Criteria
Structural validity	+	<p>CTT: CFA: CFI or TLI or comparable measure >0.95 OR RMSEA <0.06 OR SRMR <0.08²</p> <p>IRT/Rasch: No violation of <u>unidimensionality</u>³: CFI or TLI or comparable measure >0.95 OR RMSEA <0.06 OR SRMR <0.08 <i>AND</i> no violation of <u>local independence</u>: residual correlations among the items after controlling for the dominant factor < 0.20 OR Q3's < 0.37 <i>AND</i> no violation of <u>monotonicity</u>: adequate looking graphs OR item scalability >0.30 <i>AND</i> adequate <u>model fit</u>: IRT: $\chi^2 > 0.01$ Rasch: infit and outfit mean squares ≥ 0.5 and ≤ 1.5 OR Z-standardized values > -2 and <2</p>
	?	CTT: Not all information for '+' reported IRT/Rasch: Model fit not reported
	-	Criteria for '+' not met
Internal consistency	+	At least low evidence ⁴ for sufficient structural validity ⁵ AND Cronbach's alpha(s) ≥ 0.70 for each unidimensional scale or subscale ⁶
	?	Criteria for "At least low evidence ⁴ for sufficient structural validity ⁵ " not met
	-	At least low evidence ⁴ for sufficient structural validity ⁵ AND Cronbach's alpha(s) < 0.70 for each unidimensional scale or subscale ⁶

Reliability	+	ICC or weighted Kappa ≥ 0.70
	?	ICC or weighted Kappa not reported
	-	ICC or weighted Kappa < 0.70
Measurement error	+	SDC or LoA $< MIC^5$
	?	MIC not defined
	-	SDC or LoA $> MIC^5$
Hypotheses testing for construct validity	+	The result is in accordance with the hypothesis ⁷
	?	No hypothesis defined (by the review team)
	-	The result is not in accordance with the hypothesis ⁷
Cross-cultural validity\measurement invariance	+	No important differences found between group factors (such as age, gender, language) in multiple group factor analysis OR no important DIF for group factors (McFadden's $R^2 < 0.02$)
	?	No multiple group factor analysis OR DIF analysis performed
	-	Important differences between group factors OR DIF was found
Criterion validity	+	Correlation with gold standard ≥ 0.70 OR AUC ≥ 0.70
	?	Not all information for '+' reported
	-	Correlation with gold standard < 0.70 OR AUC < 0.70
Responsiveness	+	The result is in accordance with the hypothesis ⁷ OR AUC ≥ 0.70
	?	No hypothesis defined (by the review team)
	-	The result is not in accordance with the hypothesis ⁷ OR AUC < 0.70

The criteria are based on e.g. Terwee et al.(30) and Prinsen et al.(5)

AUC = area under the curve, CFA = confirmatory factor analysis, CFI = comparative fit index, CTT = classical test theory, DIF = differential item functioning, ICC = intraclass correlation coefficient, IRT = item response theory, LoA = limits of agreement, MIC = minimal important change, RMSEA: Root Mean Square Error of Approximation, SEM = Standard Error of Measurement, SDC = smallest detectable change, SRMR: Standardized Root Mean Residuals, TLI = Tucker-Lewis index

Figura 6. Criterios de Terwee y Prinsen.

4.1 PROPIEDADES PSICOMÉTRICAS: CALIDAD DE LOS RESULTADOS

4.1.1 Validez de contenido

Determina hasta qué punto los conceptos de interés están representados de forma exhaustiva por los elementos del cuestionario (3). Para poder calificar la calidad de un cuestionario, los autores deben proporcionar una descripción clara de los siguientes aspectos relacionados con el desarrollo del mismo:

- *Objetivo de la medición del cuestionario.*

Puede ser discriminativo, evaluativo o predictivo. Es importante porque diferentes elementos pueden ser válidos para diferentes objetivos (107).

- *Población objetivo (población para la que se desarrolló el cuestionario).*

Una descripción adecuada de la población objetivo es trascendente para juzgar la amplitud y la aplicabilidad del cuestionario en otras poblaciones (107).

- *Conceptos que se pretenden medir con el cuestionario.*

Para juzgar la idoneidad de un cuestionario para un propósito específico, es significativo que los autores proporcionen un marco claro sobre cuál es el concepto general que se va a medir. Los conceptos relevantes se pueden definir en términos de síntomas, funcionamiento (físico, psicológico o social), percepciones generales de salud o calidad de vida en general. Estos diferentes niveles de resultados deben distinguirse y medirse claramente mediante subescalas independientes (107).

- *Selección y reducción de artículos.*

Los métodos para la selección y reducción de ítems y la ejecución de un estudio piloto para examinar la legibilidad y comprensión deben estar justificados e informados. Los elementos del cuestionario deben reflejar áreas que son importantes para la población objetivo que se está estudiando, y por ello es importante la participación de dicha población durante la selección del artículo (107).

- *Interpretabilidad de los ítems.*

Completar el cuestionario no debería requerir habilidades de lectura más allá de las de un niño de educación primaria para evitar valores perdidos y respuestas poco fiables. Para ello, los elementos tienen que ser breves y simples, estar exentos de términos o palabras complicadas, y se debe evitar la presencia de dos preguntas de forma simultánea. Además, el periodo de tiempo al que se refieren las preguntas debe estar claramente establecido y justificado (107).

Para calificar la validez de contenido de un cuestionario hay 3 pasos a seguir:

Primero, cada resultado de un solo estudio sobre el desarrollo de un cuestionario y la validez de contenido se califica en función de los 10 criterios para una buena validez de contenido. Los revisores califican el contenido del cuestionario ellos mismos (8) (figura 7).

	PROM development study	Content validity study	Reviewers' rating	Remarks
1	The construct of interest is clearly described (i.e. 'very good' rating of box 1 standard 1), the origin of construct is clear (i.e. 'very good' rating of box 1 standard 2) and there is evidence from concept elicitation, literature, or professionals that at least 85% of the items refer to the construct of interest.	Professionals rated the relevance of the items for the construct of interest in a content validity study that was not inadequate (i.e. 'very good', 'adequate' or 'doubtful' rating for quality of relevance study in box 2d, standards 22-26) and found at least 85% of the items relevant for the construct.	Reviewers consider at least 85% of the items relevant for the construct of interest.	Every PROM item should measure a defined facet of the construct of interest, within the conceptual framework. PROM items should also be specific for the construct of interest, i.e. they should not measure a co-existing, but separate construct. For example, an item in a fatigue questionnaire such as "my muscles are weak" is relevant to fatigue but not specific for fatigue. Someone with MS may answer this question in the affirmative but not experiencing fatigue. There should also be no unnecessary items (too many items, except for a large scale item bank that will be used for computer adaptive testing). When a total PROM score is evaluated, each subscale (domain) should be relevant for the construct that the total PROM intends to measure. Professionals can best ensure that items are consistent with the theory, conceptual framework or disease model that was used to define the construct of interest.
2	The target population of interest is clearly described (i.e. 'very good' rating box 1 standard 3) and representative patients were involved in the elicitation of relevant items (i.e. 'very good' or 'adequate' rating box 1 standard 5) and concept elicitation ('worst score counts' box 1a standards 6-13) was not inadequate. If it is doubtful whether the study was performed in a sample representing the target population, we recommend giving an indeterminate (?) rating.	Patients rated the relevance of the items for them in a content validity study that was not inadequate (rating for quality of relevance study, box 2a standards 1-7) and found at least 85% of the items relevant for them.	Reviewers consider at least 85% of the items relevant for the population of interest.	The relevance of the items for the target population can best be judged by patients. Some items may be relevant to only a small number of patients but they are necessary to capture the full range of patient experiences.
3	The context of use of interest is clearly described (i.e. 'very good' rating box 1 standard 4).	Professionals rated the relevance of the items for the context of use of interest in a content validity study that was not inadequate (rating for quality of relevance study, i.e. box 2d, standards 22-26) and found at least 85% of the items relevant for the context of use.	Reviewers consider at least 85% of the items relevant for the context of use of interest.	It should especially be clear whether the PROM is suitable for use in research and/or clinical practice. Professionals are considered to be more knowledgeable about the context of use of the PROM than patients.

4	A justification is provided for the response options.	Patients or professionals rated the appropriateness of the response options in a content validity study that was not inadequate (rating for quality of relevance study, i.e. box 2a standards 1-7 or box 2d standards 22-26) and found at least 85% of the response options relevant.	Reviewers consider the at least 85% of the response options appropriate for the construct, population, and context of use of interest.	The response options should be appropriate for the construct, population, and context of use of interest. For example, if the construct is pain intensity, the response options should measure intensity, not frequency. Also, a reasonable range of responses should be provided for measuring the construct of interest.
5	A justification is provided for the recall period.	Patients or professionals rated the appropriateness of the recall period in a content validity study that was not inadequate (rating for quality of relevance study, i.e. box 2a standards 1-7 or box 2d standards 22-26) and found the recall period appropriate.	Reviewers consider the recall period appropriate for the construct, population, and context of use of interest.	The recall period can be important for measuring the construct, for example, whether there is no recall period (do you feel depressed now?) or whether the recall period is 1 week (did you feel depressed last week?). Different recall periods may be important, depending on the context. However, sometimes it does not matter whether the recall period is e.g. 1 or 2 weeks.
6	Patients were asked about the comprehensiveness of the PROM in the concept elicitation phase or in a cognitive interview study that was not inadequate (rating for quality of comprehensiveness study, i.e. box 1a standards 6-13, or box 1b standards 26-35) and no key concepts were missing.	Patients or professionals were asked about the comprehensiveness of the PROM in a content validity study that was not inadequate (rating for quality of comprehensiveness study, i.e. box 2b standards 8-14, or box 2e standards 27-31) and no key concepts were missing.	Reviewers consider the PROM comprehensive for the construct, population and context of use of interest.	The items should cover the full breadth of the construct of interest. However, there are often good reasons for not including all content suggested by patients in a PROM, for example because an item (or domain) is considered to be outside the scope of the PROM. When a total PROM score is evaluated, the subscales (domain) together should cover the full breadth of the construct that the total PROM intends to measure.
7	Patients were asked about the comprehensibility of the instructions (including recall period) in a cognitive interview study that was not inadequate (rating for quality of comprehensibility study, i.e. box 1b standards 16-25) and problems were adequately addressed.	Patients were asked about the comprehensibility of the instructions (including recall period) in a content validity study that was not inadequate (rating for quality of comprehensibility study, box 2c standards 15-21) and no important problems were found.		
8	Patients were asked about the comprehensibility of the items and response options (including wording of the items and response options) in a cognitive interview study that was not inadequate (rating for quality of comprehensibility study, box 1b standards 16-25) and problems were adequately addressed.	Patients were asked about the comprehensibility of the items and response options in a content validity study that was not inadequate (rating for quality of comprehensibility study, box 2c standards 15-21) and no important problems were found for at least 85% of the items and response options.		
9			Reviewers consider at least 85% of the items and response options appropriately worded.	Consider aspects such as reading level (a scale should not require reading skills beyond that of a 12-year old), ambiguous items, double-barrelled questions, jargon, value-laden words, and length or items [97].
10			Reviewers consider at least 85% of the response options matching the questions.	The response options should be appropriate to the question asked and should be linguistically linked to the item content.

Figura 7. Guía de los 10 criterios para una buena validez de contenido.

Segundo, los resultados de todos los estudios disponibles se resumen cualitativamente para determinar si, en general, la relevancia, la exhaustividad, la comprensibilidad y la validez general del contenido son calificables como suficiente (+), insuficiente (-), inconsistente (±) o indeterminada (?), teniendo en cuenta toda la evidencia existente.

En este caso, la atención se centra en el cuestionario, mientras que en el paso anterior, la atención se centraba en los estudios individuales.

Estas calificaciones globales se completan en la penúltima columna de la tabla y se califican como suficientes (+), insuficientes (-) o inconsistentes (±). Una calificación general indeterminada (?) en esta fase no es posible, porque la calificación del revisor siempre estará disponible: será suficiente (+) o insuficiente (-) o inconsistente (±) (8) (figura 8).

Name of the PROM or subscale:	PROM development study	Content validity study 1	Content validity study 2 ²	Rating of reviewers	OVERALL RATINGS PER PROM ³ (see step 3b)	QUALITY OF EVIDENCE (see step 3c)
Criteria	+ / - / ± / ? ¹	+ / - / ± / ?	+ / - / ± / ?	+ / - / ± / ?	+ / - / ±	High, moderate, low, very low
Relevance						
1 Are the included items relevant for the construct of interest? ⁴						
2 Are the included items relevant for the target population of interest? ⁴						
3 Are the included items relevant for the context of use of interest? ⁴						
4 Are the response options appropriate?						
5 Is the recall period appropriate?						
RELEVANCE RATING						
Comprehensiveness						
6 Are all key concepts included?						
COMPREHENSIVENESS RATING						
Comprehensibility						
7 Are the PROM instructions understood by the population of interest as intended?						
8 Are the PROM items and response options understood by the population of interest as intended?						
9 Are the PROM items appropriately worded?						
10 Do the response options match the question?						
COMPREHENSIBILITY RATING						
CONTENT VALIDITY RATING						

¹ Ratings for the 10 criteria can only be + / - / ?. The RELEVANCE, COMPREHENSIVENESS, COMPREHENSIBILITY, AND CONTENT VALIDITY ratings can be + / - / ± / ?

² Add more columns if more content validity studies are available

³ If ratings are inconsistent between studies, consider using separate tables for subgroups of studies with consistent results.

⁴ These criteria refer to the construct, population, and context of use of interest in the systematic review.

Figura 8. Criterios COSMIN y sistema de calificación para evaluar la validez de contenido.

Tercero, las calificaciones globales establecidas en el paso anterior irán acompañadas de una calificación de la calidad de la evidencia, utilizando el método GRADE modificado (alto, moderado, bajo o muy bajo), que nos dará la certeza necesaria para corroborar que esas calificaciones generales son fiables (8) (figura 9). Este método está destinado a degradar la evidencia cuando existen inquietudes sobre la calidad de la misma. El punto de partida es siempre la suposición de que la calificación global es de alta calidad, para posteriormente degradar esa calidad de la evidencia en uno o dos niveles por factor; a moderada, baja o muy baja cuando la calidad de los estudios es dudosa o inadecuada (existe riesgo de sesgo), o cuando hay inconsistencia (es inexplicable) o resultados indirectos.

En este método GRADE modificado no se tienen en cuenta ni la imprecisión ni el sesgo de publicación, ya que la imprecisión es menos relevante para la validez de contenido porque el desarrollo del cuestionario y los estudios de validez de contenido se refieren a un tipo de investigación cualitativa, mientras que el sesgo de publicación es difícil de evaluar debido a la falta de registros de estudios para el desarrollo de la nueva herramienta y a la carencia de investigaciones sobre validez de contenido (8).

Study design	Quality of evidence	Lower if
At least 1 content validity study	High	Risk of bias
No content validity studies	Moderate	-1 Serious
	Low	-2 Very serious
	Very low	-3 Very serious
		Inconsistency
		-1 Serious
		-2 Very serious
		Indirectness
		-1 Serious
		-2 Very serious

Figura 9. Clasificación de la calidad de la evidencia sobre la validez de contenido (método GRADE modificado).

Según los criterios de Terwee y Prinse, para obtener una calificación suficiente para la validez de contenido de un cuestionario, es recomendable reportar la siguiente información en una revisión sistemática: características de los cuestionarios (constructo, población objetivo, contexto de uso previsto, número de escalas, número de ítems, período de recuperación, etc), calidad del desarrollo del cuestionario, calidad y características de los estudios de validez de contenido disponibles (población del estudio, tamaño de la muestra, diseño del estudio, pacientes y/o profesionales involucrados), calidad asociada de la evidencia, calificación de la relevancia, exhaustividad, comprensibilidad, y validez general de contenido por cuestionario.

En ausencia de una descripción clara, la validez de contenido se considera inconsistente (8).

4.1.2 Validez estructural

Se trata del grado en que las puntuaciones de un instrumento son un reflejo adecuado de una dimensión del constructo a medir. Generalmente, se evalúa mediante análisis factorial o análisis IRT (teoría de respuesta al ítem) / Rasch (3).

En una revisión sistemática es útil hacer una distinción entre aquellos estudios en los que se realiza un análisis factorial para evaluar la validez estructural y aquellos en los que se emplea dicho análisis para evaluar la unidimensionalidad de cada subescala por separado o por subescala.

Para medir la validez estructural, el EFA se efectúa para cada uno de sus elementos evaluando el número hipotético de subescalas de la herramienta y la agrupación de elementos dentro de las subescalas (estudios de validez estructural). Por otra parte, para medir la unidimensionalidad por subescala se efectúan análisis de factores múltiples en los ítems de cada subescala por separado, para evaluar si cada subescala en su propia medida evalúa un solo constructo (estudios de unidimensionalidad) (5).

Para que la puntuación de esta propiedad psicométrica pueda ser considerada como suficiente, deben darse los siguientes supuestos: CTT (teoría clásica de los tests) (CFA: CFI o TLI o medida comparable > 0.95 o RMSEA < 0.06 o SRMR < 0.082); IRT / Rasch (Sin violación de la unidimensionalidad: CFI o TLI o

medida comparable > 0.95 o $RMSEA < 0.06$ o $SRMR < 0.08$), y sin violación de la independencia local: correlaciones residuales entre los ítems después de controlar el factor dominante < 0.20 o $Q3's < 0.37$ (5,7).

4.1.3 Consistencia interna

Medida del grado en que los ítems en una escala o subescala del cuestionario están correlacionados, midiendo así un mismo concepto (3). Se trata de una propiedad de medición importante para los cuestionarios que pretenden evaluar un único constructo mediante el uso de varios elementos.

Una escala internamente consistente (homogénea o unidimensional) se consigue mediante una definición de constructo apropiada, los ítems idóneos, un EFA y un CFA (108).

El método estadístico más utilizado para medir en un instrumento esa consistencia interna, es el alfa de Cronbach. Un alfa de Cronbach bajo muestra una falta de correlación entre los elementos de una escala, lo que hace que el resumen de los elementos sea injustificado. Por el contrario, un alfa de Cronbach muy alto (valores próximos a 1) indica correlaciones altas entre los elementos de la escala, es decir, redundancia o duplicación en relación con los ítems (71).

Terwee y Prinsen conceden una calificación de suficiente para dicha propiedad psicométrica cuando existe al menos baja evidencia de suficiente validez estructural y un alfa de Cronbach ≥ 0.70 para cada escala o subescala unidimensional (5,7).

4.1.4 Fiabilidad

Definimos fiabilidad como la estabilidad manifiesta de una herramienta en mediciones repetidas (3). Desde el punto de vista estadístico, el ICC es el índice más utilizado y adecuado para ponderar la concordancia entre distintas mediciones de una variable, y este hace referencia a la variación en la población dividida por la variación total (variación inter-individual (variación de la población) más variación intra-individual (error de medición)), expresada como una relación entre 0 y 1 (35,68).

El período de tiempo transcurrido entre la repetición de las administraciones debe ser lo suficientemente largo para evitar la recuperación y lo suficientemente corto para garantizar que no se haya producido un cambio clínico. Una o dos semanas serán idóneas, pero podría haber razones para pensar lo contrario, como cuando se administra una intervención en el período intermedio y muchos de esos pacientes sufren cambios, cuando influye el entorno en el que se administró el instrumento, cuando las instrucciones dadas para completarlo no están claras o son insuficientes, etc (7,48,107).

Por lo tanto, Terwee y Prinse no califican la idoneidad del período de tiempo, sino que este período de tiempo esté descrito y justificado, y otorgan una calificación suficiente cuando ICC o Kappa ponderado ≥ 0.70 (5,7).

4.1.5 Error de medición

Error sistemático y aleatorio en la puntuación de un paciente que no se atribuye a cambios reales en los fenómenos que se están midiendo (3). Para aplicar los criterios de un buen error de medición se necesita información sobre el cambio detectable más pequeño (SDC) o los límites de acuerdo (LoA), además del cambio mínimo importante (MIC).

El MIC se calcula mejor a partir de múltiples estudios y utilizando múltiples anclajes (109–112). Si no hay suficiente información disponible para juzgar si el SDC o LoA es más pequeño que el MIC, se recomienda simplemente reportar la información que está disponible en el SDC o LoA, sin calificar la calidad de la evidencia.

En Terwee y Prinse la calificación se considera suficiente cuando SDC o LoA $< MIC$ (esta evidencia puede provenir de diferentes estudios) (5,7).

4.1.6 Prueba de hipótesis para la validez de constructo

Grado en que las puntuaciones de un instrumento en particular se relacionan con otras medidas de una manera que es consistente con hipótesis derivadas teóricamente sobre los conceptos que se están midiendo (3). La validez de constructo debe evaluarse probando hipótesis predefinidas, y estas deben ser lo más específicas posibles. Sin hipótesis específicas el riesgo de sesgo es alto, ya

que de manera retrospectiva se pueden facilitar de una forma errónea explicaciones alternativas para las correlaciones bajas en lugar de concluir que el cuestionario puede no ser válido (107).

Terwee y Prinsen dan una calificación suficiente para la validez de constructo si las hipótesis se especifican de antemano y si al menos el 75% de los resultados están en correspondencia con estas hipótesis (5,7).

4.1.7 Validez transcultural / invariancia de medición

Es el grado en que el desempeño de los elementos en un instrumento traducido o adaptado culturalmente es un reflejo adecuado de su versión original (3). Se debe evaluar cuando un instrumento se usa o se va a usar en diferentes poblaciones *culturales* (difieren en etnia, idioma, género o grupos de edad) (7), y esto se consigue midiendo el funcionamiento diferencial de los ítems (DIF) mediante el análisis de regresión logística o el CFA multigrupo (MGCFA), si es que la estructura factorial y las cargas factoriales son equivalentes entre los grupos.

La invariancia de medición y la no DIF hacen referencia a si los encuestados de diferentes grupos con el mismo nivel de rasgo latente (teniendo en cuenta las diferencias de grupo) responden de manera similar a un elemento en particular (113). Terwee y Prinse otorgan una calificación suficiente para esta propiedad psicométrica siempre que no se encuentren diferencias importantes entre los factores del grupo (como la edad, el género, el idioma) en el análisis de factores de grupos múltiples o cuando no se encuentren DIF significativos para los factores del grupo (R^2 de McFadden's < 0.02) (5,7).

4.1.8 Validez de criterio

Es la medida en que las puntuaciones de un instrumento en particular se relacionan con un gold standard (3). Terwee y Prinsen atribuyen una calificación suficiente para la validez de criterio si la correlación con el gold standard es ≥ 0.70 o el área bajo la curva (AUC) ≥ 0.70 (5,7).

4.1.9 Sensibilidad al cambio

Es la capacidad de un cuestionario para detectar cambios clínicamente importantes a lo largo del tiempo, incluso si estos cambios son mínimos (3). Se trata de una medida de validez longitudinal y, por lo tanto, se evalúa probando hipótesis predefinidas (esto muestra la capacidad de un cuestionario para medir cambios si realmente han ocurrido) (114).

Además, el instrumento debería poder distinguir entre cambios clínicamente importantes y errores de medición, de modo que la capacidad de respuesta debe probarse relacionando el cambio mínimo detectable (SDC) con el MIC o a través del área bajo la curva de características operativas del receptor (AUC/ROC) (115), método que permite en un cuestionario distinguir pacientes que han cambiado y no han cambiado, según un criterio externo.

Terwee y Prinse califican como suficiente un $AUC \geq 0.70$ y al menos el 75% de los resultados están de acuerdo con las hipótesis (5,7).

Método GRADE (calificación de recomendaciones, valoración, desarrollo y evaluaciones)

5. MÉTODO GRADE

GRADE (calificación de recomendaciones, valoración, desarrollo y evaluaciones) es un marco transparente para desarrollar, presentar resúmenes de evidencia y proporcionar un enfoque sistemático para hacer recomendaciones de práctica clínica (116–118). Se trata de un método utilizado por revisores sistemáticos y desarrolladores de guías para evaluar la calidad de la evidencia y decidir si recomendar una intervención (117). GRADE difiere de otras herramientas de evaluación fundamentalmente por 3 razones: separa la calidad de la evidencia y la fuerza de la recomendación, la calidad de la evidencia se evalúa para cada resultado, y en los estudios observacionales se puede “aumentar la calidad de su evidencia” si estos cumplen con ciertos criterios.

GRADE está emergiendo como el método dominante para evaluar estudios controlados y hacer recomendaciones para revisiones sistemáticas y guías (117,119–128). Por lo general, para comprender y poder hacer uso de este método se requiere leer una serie de publicaciones destinadas a ello (119–128), y quizá asistir a un taller para alcanzar los suficientes conocimientos y poder llevar a cabo, *a posteriori*, una evaluación crítica. No existen explicaciones concisas sobre GRADE destinadas a una audiencia de investigadores en general, ni libros de texto alguno donde se incluya.

Además, debemos de tener en cuenta que dicho método es subjetivo. No se puede implementar mecánicamente, es necesario que exista una cantidad considerable de subjetividad en cada decisión, por lo que dos personas que evalúan el mismo conjunto de pruebas podrían llegar razonablemente a conclusiones diferentes sobre su certeza. Desde otra perspectiva, este método lo que sí proporciona es un marco reproducible y transparente para calificar la certeza en la evidencia (129).

5.1 EVALUACIÓN DE LA CALIDAD DE LA EVIDENCIA Y ELABORACIÓN DE LA RECOMENDACIÓN.

El método GRADE podría explicarse a través de estos 5 puntos:

En el primer punto, tenemos que tener en cuenta que GRADE tiene cuatro niveles de evidencia, llamados también certeza en la evidencia o calidad de la evidencia: “muy bajo”, “bajo”, “moderado” y “alto” (figura 10). Debemos asignar una calificación *a priori* de "alto" para los ensayos clínicos aleatorizados (ECAs) y "bajo" para los estudios observacionales. A los ECAs se les asigna inicialmente una calificación más alta porque generalmente son menos propensos al sesgo que los estudios observacionales (130,131).

Quality level	Definition
High	We are very confident that the true measurement property lies close to that of the estimate* of the measurement property
Moderate	We are moderately confident in the measurement property estimate: the true measurement property is likely to be close to the estimate of the measurement property, but there is a possibility that it is substantially different
Low	Our confidence in the measurement property estimate is limited: the true measurement property may be substantially different from the estimate of the measurement property
Very low	We have very little confidence in the measurement property estimate: the true measurement property is likely to be substantially different from the estimate of the measurement property

Figura 10. Definición de los niveles de la calidad de la evidencia en el método GRADE.

En un segundo punto, se lleva a cabo una clasificación inicial para determinar la “disminución” o el “aumento” del nivel de certeza. Es común que los ECAs y los estudios observacionales sufran una disminución del nivel de certeza porque adolecen de sesgos identificables. Del mismo modo, los estudios observacionales pueden aumentar su nivel de certeza cuando múltiples estudios de alta calidad muestran resultados consistentes (132).

- **Factores que disminuyen la calidad de la evidencia**

Los dominios presentes en GRADE que originan esa disminución de la calidad de la evidencia son: riesgo de sesgo, inconsistencia, ausencia de evidencia directa, imprecisión y sesgo de publicación. Para cada uno de estos, los autores tienen la opción de disminuir su nivel de certeza uno o dos niveles (103) (figura 11).

- ***Riesgo de sesgo (limitaciones en el diseño y en la ejecución)***

El sesgo ocurre cuando los resultados de un estudio no representan la realidad debido a limitaciones inherentes en el diseño y/o en la realización de este, o siempre que la calidad de los estudios adicionales de validez de contenido sea dudosa (126).

En la práctica, es difícil saber hasta qué punto los sesgos potenciales influyen en los resultados. GRADE se utiliza para calificar el conjunto de pruebas en cuanto a resultado en lugar de en cuanto a estudios, por lo que los autores deben juzgar si el riesgo de sesgo en los estudios individuales es lo suficientemente grande como para que su confianza en el efecto estimado del tratamiento sea menor (126). Algunos de los factores que pueden llevar al sesgo son: ausencia de una secuencia de asignación claramente aleatorizada, falta de cegamiento, carencia de ocultación en la asignación, anteponer los tratamientos a las pruebas, interrupción de la investigación o grandes pérdidas durante el seguimiento.

- ***Inconsistencia de los resultados***

Se produce cuando existe una variabilidad significativa e inexplicable en los resultados de diferentes ensayos. La calidad de la evidencia es más alta cuando se presentan varios estudios que muestran efectos consistentes. Al considerar si la certeza debe calificarse o no por inconsistencia, los autores deben inspeccionar la similitud de las estimaciones puntuales y la superposición de sus intervalos de confianza, así como los criterios estadísticos de heterogeneidad (123).

- *Ausencia de evidencia directa*

Existe una mayor evidencia cuando los estudios comparan directamente las intervenciones de interés en la población objetivo e informan de los resultados críticos para la toma de decisiones. La ausencia de evidencia directa va a aparecer cuando existan diferencias en la población (aplicabilidad), en las intervenciones (aplicabilidad), en las medidas de resultado (resultados sustitutos) o se establezcan comparaciones indirectas (cuando no tenemos comparaciones directas entre dos o más intervenciones de interés) (127).

- *Imprecisión*

Se produce cuando la amplitud en los intervalos de confianza afecta a la calidad de los datos. El método GRADE para calificar la imprecisión se centra en el intervalo de confianza del 95% en torno a la mejor estimación del efecto absoluto y el número de sujetos o eventos incluidos (122). La imprecisión es menos relevante para la validez de contenido porque el desarrollo de una nueva herramienta y los estudios de validez de contenido se encuadran dentro de la investigación cualitativa (8).

- *Sesgo de publicación*

El sesgo de publicación es quizá el más incómodo de los dominios GRADE porque requiere hacer inferencias sobre la evidencia faltante. Por tanto, es el más difícil de evaluar debido a la falta de registros acerca de estudios sobre el desarrollo de esa nueva herramienta y estudios sobre validez de contenido (8). Es más habitual en presencia de un conjunto de ensayos de reducido tamaño, positivos y financiados por parte de la industria (124).

• **Factores que aumentan la calidad de la evidencia**

Al igual que existen una serie de factores que disminuyen la calidad de la evidencia, también existen una serie de condicionantes que la aumentan. En primer lugar, lo que se conoce como “fuerza de asociación”, que se produce cuando el efecto es tan grande que el sesgo común a los estudios observacionales no puede explicar el resultado. En segundo lugar, el “gradiente dosis-respuesta”,

que se da cuando el resultado es proporcional al grado de exposición. En tercer lugar, cuando surgen “factores de confusión que solo reducen el tamaño del efecto”, produciéndose en la mayoría de las ocasiones una disminución de la confusión residual en lugar de un aumento de la magnitud del efecto (128) (figura 11).

Study Design	Quality of Evidence	Lower if	Higher if
Randomized trial →	High	Risk of bias -1 Serious -2 Very serious	Large effect +1 Large +2 Very large
	Moderate	Inconsistency -1 Serious -2 Very serious	Dose response +1 Evidence of a gradient
Observational study →	Low	Indirectness -1 Serious -2 Very serious	All plausible confounding +1 Would reduce a demonstrated effect or
	Very low	Imprecision -1 Serious -2 Very serious Publication bias -1 Likely -2 Very likely	+1 Would suggest a spurious effect when results show no effect

Figura 11. Factores que disminuyen y aumentan la calidad de la evidencia en GRADE.

En un tercer punto, se asigna una calificación final a la calidad de la evidencia de "alta", "moderada", "baja" o "muy baja" para todos los resultados de importancia crítica (116). Para ello, se va aumentando y disminuyendo la calidad de la evidencia mediante un juicio y una reflexión cuidadosa. Una regla general es subir o bajar una categoría por tema, de tal modo que la calidad de la evidencia para un resultado estudiado en un ECA que inicialmente comienza siendo “alta”, podría pasar a “moderada” debido al alto riesgo de sesgo en los estudios incluidos; descender más hasta “baja”, si hubo una cantidad significativa de heterogeneidad inexplicable entre los ensayos; e incluso más aún, hasta “muy

baja", si existió un número reducido de eventos que condujeran a intervalos de confianza que incluían beneficios y daños importantes (122–124,126,127).

En un cuarto punto, debemos tener en cuenta otros factores que influyen en el método o fenómeno de la recomendación. Una evidencia de alta calidad no siempre implica una recomendación sólida. Las recomendaciones deben considerar algunos otros factores además de la calidad de la evidencia, como el equilibrio entre efectos deseables e indeseables, la calidad global de la evidencia, los valores y preferencias de los pacientes en los casos en que la relación beneficio / riesgo es menos clara y, en último lugar, los recursos / costes (119).

Finalmente, **en el quinto punto**, hablamos de hacer recomendaciones fuertes o débiles y a favor o en contra de una intervención, teniendo en cuenta los 4 factores comentados en el punto anterior (equilibrio efectos deseables / indeseables, calidad global de la evidencia, valores y preferencias de los pacientes en la relación beneficio / riesgo, y recursos / costes) (133). Las recomendaciones fuertes sugieren que todas o casi todas las personas elegirían esa intervención, mientras que las recomendaciones débiles implican que es posible que exista una variación importante en la decisión que probablemente tomen las personas informadas. La fuerza de las recomendaciones es factible: una recomendación débil indica que participar en un proceso de toma de decisiones compartido es esencial, mientras que una recomendación fuerte propone que generalmente no es necesario presentar ambas opciones.

Es más probable que las recomendaciones sean débiles en lugar de fuertes cuando la certeza de la evidencia es baja y existe un equilibrio cercano entre las consecuencias deseables e indeseables, cuando hay una variación sustancial o incertidumbre en los valores y preferencias de los pacientes, y cuando las intervenciones requieren recursos considerables (117,133–135). Con relación a la forma de redactar las recomendaciones, GRADE plantea para las fuertes emplear el término *se recomienda* y para las débiles hacer uso del término *se sugiere* (133).

Lista de verificación STARD

6. LISTA DE VERIFICACIÓN STARD

La STARD (STAndards for the Reporting of Diagnostic accuracy studies) checklist se creó con el objetivo de perfeccionar el diseño sobre los estudios de precisión acerca de pruebas diagnósticas. El grupo de trabajo de la Cochrane se reunió en Roma en 1999, continuando el ejemplo de la iniciativa CONSORT (CONsolidated Standards of Reporting Trials) que valora los ensayos clínicos (136,137). Tras analizar las publicaciones y principales bases de datos, el grupo de trabajo extrajo 33 listas a partir de las cuales se incluyeron 75 puntos potenciales para posteriormente confeccionar un documento en el que participaron investigadores, editores, metodólogos y organizaciones profesionales (miembros del propio grupo de trabajo de STARD) (138).

El resultado final fue un listado de 25 ítems divididos en 5 dominios: título/resumen/palabras clave, introducción, métodos, resultados, discusión (101,139), y un diagrama de flujo en el que se incluían los pasos necesarios a seguir: orden de ejecución, criterios de inclusión de los pacientes, prueba de referencia y número de individuos sometidos a dicha prueba (138) (figura 12).

La lista de verificación STARD tiene como objetivo principal facilitar un listado que sirva de guía para la publicación de los estudios de precisión de pruebas diagnósticas. Está destinada a editores de revistas y autores de artículos para evaluar los sesgos potenciales de la investigación, validez interna y la aplicabilidad o generalización de los resultados (validez externa) (138).

Los resultados fueron publicados en algunas revistas médicas importantes en el año 2003 (101). A raíz de esto, muchas editoriales encomendaron a los autores y a los editores su aceptación, ya que la incorporación de STARD optimizaba el proceso de evaluación de la investigación (140). Su uso fue promovido por el Comité Internacional de Editores de Revistas Médicas, que en el año 2006 implantó su manejo como requisito indispensable para la publicación de los artículos (141).

Su aprobación no tuvo la aceptación deseada y, en la actualidad, la lista de verificación STARD se emplea en un mínimo porcentaje de publicaciones (142).

Esto no es indicativo de que estos estudios sean de mala calidad, sino que es preciso incluir una descripción del proceso para que se puedan explicar con una mayor transparencia los resultados (137). En España, en algunos estudios sobre publicaciones biomédicas, se concluye que prácticamente menos de la mitad de los artículos evaluados siguen los pasos de la lista de verificación STARD (143).

6.1 ÍTEMS

6.1.1 Título / resumen / palabras clave

El título nos debe indicar que la investigación versa sobre precisión de pruebas diagnósticas. Se aconseja también incluir entre las palabras clave los términos *especificidad* y *sensibilidad*, aunque esto no es garantía de que el estudio sea adecuadamente reconocido como tal, por lo que el resumen del artículo también debe ser lo más claro posible.

6.1.2 Introducción

Para permitir que el lector tenga una idea lo más lúcida posible sobre el tema de estudio, se debe exponer de una manera clara qué es lo que se está evaluando y qué tipo de pacientes o condiciones se incluyen en el estudio, para dar así respuesta a la pregunta u objetivos de la investigación.

6.1.3 Métodos

- ***Población de estudio:*** deben ser descritos los criterios de inclusión y exclusión que se emplearon, y el lugar y las condiciones donde fueron obtenidos dichos datos. Es imprescindible que dichas descripciones sean lo más precisas posibles, ya que posibilitarán al lector a establecer si estos resultados son adaptables a sus circunstancias.

- ***Reclutamiento de los pacientes:*** se detalla la forma por la cual los pacientes fueron seleccionados para el estudio (síntomas, pruebas previas, test de referencia, sexo, edad, idioma, etc).

-
- **Obtención de la muestra:** hay que especificar la inclusión en el estudio de aquellos participantes que cumplieron con los criterios del punto 3 y 4 (población de estudio y reclutamiento de los pacientes), y si fueron incluidos de manera consecutiva o a través de algún tipo de muestreo (en caso de haber existido, este debe ser puntualizado).
 - **Recolección de datos:** hay que señalar si la recopilación de los datos fue proyectada después de llevar a cabo los test (investigación retrospectiva) o bien antes de aplicarlos (investigación prospectiva), ya que esto nos va a posibilitar para distinguir la presencia de sesgos y dificultades propias de uno u otro tipo de estudios.
 - **Prueba de referencia y su justificación:** se deben detallar las pruebas empleadas como referencia y razonar el por qué son apropiadas para ese determinado tipo de estudio. Si se utilizan pruebas múltiples o de seguimiento deben especificarse y argumentarse, tanto en su conjunto como por separado.
 - **Especificaciones técnicas:** se tienen que identificar los materiales y métodos aplicados para la ejecución de las pruebas de estudio y pruebas de referencia. Igualmente, pueden facilitarse reseñas donde se detalle de manera exhaustiva la forma en que los test son administrados.
 - **Definición y coherencia de la medición:** se debe numerar y razonar el empleo de las unidades de medida, puntos de corte o categorías resultantes de las pruebas de referencia y prueba de estudio, detallándose si estas variables fueron descritas previa o posteriormente a la consecución de los resultados.
 - **Práctica de los evaluadores:** señala la preparación y experiencia de las personas que realizan e interpretan la prueba de referencia y la prueba de estudio.
 - **Enmascaramiento de los evaluadores:** debemos conocer si las personas encargadas de ejecutar e interpretar las pruebas desconocían los resultados, ya que el conocimiento de los mismos puede inducir a un sesgo diagnóstico.
 - **Métodos estadísticos:** se deben detallar los métodos, tanto para calcular o comparar las medidas de precisión diagnóstica como para cuantificar la incertidumbre diagnóstica (intervalos de confianza al 95%).

- **Cálculo de reproducibilidad:** se debe indicar si se ejecutaron pruebas para establecer la estabilidad temporal o la fiabilidad intra-observador, es decir, el grado de correspondencia confirmado al repetir una medición en condiciones similares por un mismo observador, al igual que todos los métodos estadísticos relacionados.

6.1.4 Resultados

- **Marco temporal:** se indica cuando se llevó a cabo el estudio, la inclusión de los participantes y fechas de inicio y finalización de la recolección de datos, dado que las pruebas están en continuo desarrollo.

- **Identificación de la población:** se enumeran las características clínicas y sociodemográficas de la población de estudio (edad, sexo, síntomas presentes, tratamientos, comorbilidad).

- **Inclusión de los participantes:** se detallan todos aquellos pacientes que reunieron los criterios de selección y a los que se les efectuaron pruebas de estudio y/o de referencia, además de las causas por las cuales no se les realizaron. Es importante adjuntar un diagrama de flujo donde quede plasmado todo el proceso de inclusión de los participantes.

- **Intervalo entre las distintas pruebas:** tenemos que saber el tiempo transcurrido entre la ejecución de la prueba de referencia y la prueba de estudio, y si ha sido administrado algún tratamiento entre medias.

- **Distribución de la gravedad de la enfermedad:** debemos especificar la variedad de patologías presentes en los pacientes seleccionados para el estudio, tanto en los pacientes con la condición en estudio como quienes no la ostentaban.

- **Tabulación cruzada referencia-estudio:** se debe desarrollar un cruce de frecuencias con los resultados de las pruebas de referencia y las pruebas de estudio, indicando presencia y ausencia, y resultados indeterminados y ausentes. En el caso de pruebas de estudio con medidas continuas es aconsejable exponer la distribución de los resultados para los valores de la prueba de referencia.

- *Efectos adversos*: incluye los efectos adversos que tuvieron lugar durante el estudio. Es necesario exponer la existencia de cualquier hecho que cuestione la seguridad de la prueba (invasividad o riesgos asociados a la aplicación de la prueba de estudio).

- *Estimaciones de precisión diagnóstica*: consiste en mostrar la valoración de la incertidumbre estadística incluyendo curvas ROC, medidas de especificidad, sensibilidad, valores predictivos y likelihood ratios, con sus respectivos intervalos de confianza al 95%.

- *Datos indeterminados, perdidos y valores atípicos*: es obligatorio indicar de qué forma fueron tratados dichos resultados, mostrando también diferencias entre ambos, para la prueba de referencia y la prueba de estudio.

- *Estimación de la variabilidad*: se deben detallar las posibles alteraciones en la precisión diagnóstica para los diferentes subgrupos, evaluadores y centros incluidos en la investigación. Es aconsejable que estos análisis sean planeados antes de la elección de los datos.

- *Estimación de la reproducibilidad (fiabilidad)*: se tienen que incluir los resultados de fiabilidad intra-observador o de estabilidad temporal en caso de que estos fueran realizados. Se aconseja precisar el coeficiente de variación y mostrar los factores que afectan a la reproducibilidad.

6.1.5 Discusión

- *Aplicabilidad clínica*: se aconseja a los autores discutir acerca de la validez interna y externa de los resultados, así como de las limitaciones metodológicas de dicha investigación.

SECCIÓN Y TEMA		
TÍTULO/RESUMEN/ PALABRAS CLAVE	1	Identifica el artículo como un estudio de precisión diagnóstica (se recomienda el término MeSH "sensibilidad y especificidad").
INTRODUCCIÓN	2	Especifica que las preguntas del estudio, o el objetivo del mismo, es estimar la precisión diagnóstica o comparar la precisión entre diferentes grupos y con otros test.
MÉTODOS		
Participantes	3	Describe la población incluida en el estudio: los criterios de inclusión y exclusión, lugar y localización dónde se realizó la recogida de datos.
	4	Describe cómo se seleccionaron los participantes: ¿la selección se basó en los síntomas iniciales, en los resultados de las pruebas previas, o en el hecho de que los participantes recibiesen la prueba de estudio y la prueba de referencia?
	5	Describe la selección de los participantes (el muestreo): ¿la población del estudio fue una serie consecutiva de participantes definida por los criterios de selección de los puntos 3 y 4? Si no es así, especifica cómo se seleccionaron los participantes.
	6	Describe la recogida de datos : si se planeó previamente antes de realizar la prueba de estudio y la de referencia (estudio prospectivo) o después (estudio retrospectivo).
	7	Describe la prueba de referencia y en que se basa (su justificación).
Métodos de las pruebas	8	Describe las cuestiones técnicas del material y métodos, incluyendo cómo y cuándo se tomaron las medidas, y/o cita referencias para la prueba en estudio y la prueba de referencia.
	9	Describe la definición y bases de las unidades, puntos de corte y/o categorías de los resultados de las pruebas en estudio y de la prueba de referencia.
	10	Describe el número, preparación y grado de conocimiento de las personas que realizan e interpretan la prueba de estudio y la de referencia.
	11	Describe si la aplicación de la prueba y del test de referencia se realizó de forma ciega.
Métodos estadísticos	12	Describe los métodos para calcular o comparar las medidas de precisión diagnóstica, y los métodos estadística utilizados para cuantificar la incertidumbre (por ejemplo los intervalos de confianza del 95%).
	13	Describe los métodos para calcular la reproductibilidad de la prueba, si se realizaron (es decir su fiabilidad).
RESULTADOS		
Participantes	14	Describe cuándo se realizó un estudio, incluyendo las fechas de inicio y finalización de inclusión de los participantes.
	15	Describe las características clínicas y demográficas de la población de estudio (por ejemplo edad, sexo, espectro de los síntomas, comorbilidad, tratamientos actuales, centros de reclutamiento).
	16	Describe el número de participantes que cumplen los criterios de inclusión que recibieron o no las pruebas de estudio y/o la prueba de referencia; describe por qué los participantes fallaron en recibir alguno de las pruebas (se recomienda un diagrama de flujo que identifique las pérdidas).
Resultados de las pruebas	17	Describe el intervalo de tiempo entre las pruebas de estudio y la prueba de referencia, y si se ha administrado algún tratamiento entre medias.
	18	Describe la distribución de la gravedad de la enfermedad (criterios definidos) en aquellos con la enfermedad en estudio; otros diagnósticos en los participantes sin la enfermedad en estudio.
	19	Describe los resultados en una tabla (incluyendo los indeterminados y los no obtenidos) con los resultados de la prueba de estudio y la de referencia. En las variables continuas, la distribución de los resultados según los valores de la prueba de referencia.
	20	Describe cualquier efecto adverso que pueda haber originado el realizar la prueba de estudio o la de referencia.
Estimaciones	21	Describe las estimaciones de precisión diagnóstica y las medidas de incertidumbre estadística (por ejemplo el intervalo de confianza del 95%).
	22	Describe como se interpretaron los resultados indeterminados, las pérdidas y valores atípicos o extremos de la prueba de estudio.
	23	Describe la estimación de la variabilidad de la precisión diagnóstica entre subgrupos de participantes, investigadores o centros, si se hizo.
	24	Si procede, describe la estimación de la reproductibilidad (fiabilidad) de la prueba.
COMENTARIO (DISCUSIÓN)	25	Comenta la aplicabilidad clínica de los resultados del estudio.

Figura 12. Lista de verificación STARD.

Lista de verificación STROBE

7. LISTA DE VERIFICACIÓN STROBE

En el año 2004, un grupo de estadísticos, epidemiólogos, investigadores y editores se reunió con la intención de perfeccionar la presentación de informes de alta calidad sobre estudios observacionales, basándose en STROBE (STrengthening the Reporting of Observational Studies in Epidemiology) (144) (figura 13) y tomando como referente la experiencia CONSORT (145).

El objetivo de las directrices STROBE es proporcionar una lista de verificación fácilmente entendible para garantizar una presentación clara de lo que se planificó y realizó en un estudio observacional; fomentar la producción de artículos interesantes y narrativos, manteniendo su transparencia; evaluar las fortalezas y debilidades, y facilitar a editores, revisores y lectores una interpretación y apreciación crítica de dicho artículo (144,146,147). Existen varias listas de verificación STROBE, dependiendo del tipo de estudio (una para estudios de cohortes, una para casos y controles y otra para estudios transversales) (148,149).

STROBE se fundamenta en una lista de verificación de 22 puntos y 6 dominios, la cual mantiene relación en un artículo con todas y cada una de sus secciones (título, resumen, introducción, metodología, resultados y discusión). De esos 22 puntos, 18 son universales a los 3 diseños de estudio que pueden ser evaluados por la declaración STROBE (cohortes, transversales, y casos y controles), mientras que los otros 4 puntos (ítems 6, 12, 14 y 15) están delimitados para cada uno de esos 3 tipos de estudio (150).

Como consecuencia del beneficio que reporta la aplicación de dicha lista de verificación, la cual garantiza que un artículo científico cumpla con todos los estándares en cuanto a las secciones que debe contener, es fundamental tener un conocimiento amplio sobre este prototipo de guía de comprobación de estudios observacionales y, en concreto, en aquellos autores que deseen realizar y publicar un estudio de esa misma tipología. Estas pautas se han establecido como una guía de apoyo para los investigadores, para ayudarles a construir un manuscrito

debidamente presentable que permita al lector comprenderlo y evaluarlo críticamente (150,151).

Como se ha comentado anteriormente, existen varias listas de verificación STROBE en función de la tipología del estudio. La empleada para evaluar la calidad metodológica de los artículos de una de las revisiones sistemáticas incluida en la tesis "Cancer-related fatigue measures in palliative care: a psychometric systematic review" fue la específica para estudios observacionales transversales (152).

7.1 ÍTEMS

7.1.1 Título y Resumen. Ítem 1

El diseño del estudio adoptado debe formar parte del título del manuscrito para garantizar la correcta indexación de este en las bases de datos electrónicas, asegurar la visibilidad del trabajo y aumentar el potencial de citas del mismo, con el fin de mejorar las métricas de investigación del autor, aumentar el prestigioso reconocimiento del mismo y su trabajo dentro de la comunidad científica. Este apartado debe incluir un breve resumen del estudio y presentar solo la información que se encuentre dentro del cuerpo real del manuscrito (153).

7.1.2 Introducción. Ítems 2 y 3

La introducción debe incluir información de antecedentes que determinarán el escenario para el estudio y el objetivo del mismo. El objetivo deberá indicar las intenciones del investigador para realizar el estudio y las posibles hipótesis que puedan surgir de dicho trabajo.

7.1.3 Metodología. Ítems 4-12

La metodología empleada por el investigador debe facilitar una descripción clara del diseño del estudio en una etapa temprana. Esto permitirá al lector comprender la base del estudio y ser capaz de evaluar críticamente la metodología de este. Las pautas de STROBE no permiten el uso de las palabras

prospectivo o retrospectivo, concurrente o histórico, sino que alientan al investigador a describir la metodología real (144).

Para mejorar la comprensión del lector sobre los resultados del estudio, la información sobre las herramientas de medición, entorno y ubicación deben estar presentes. A su vez, los detalles sobre el reclutamiento de la muestra variarán según el tipo de diseño observacional que se lleve a cabo, por lo que es imprescindible que el investigador conozca el diseño metodológico epidemiológico para los distintos tipos de diseño observacional existentes.

Todas las variables consideradas para el análisis descriptivo y estadístico del estudio deben anotarse dentro del apartado de métodos. Esto también incluye la notificación de los puntos de corte específicos implementados durante el análisis. Es esencial que cualquier exposición, confusión o medición de resultados se tenga en cuenta y se informe, para que el lector evalúe críticamente la fiabilidad y validez del estudio. También deben anotarse los criterios y métodos de inclusión y exclusión para superar cualquier sesgo potencial.

El método utilizado para establecer el tamaño de la muestra del estudio debe informarse junto con los intervalos de confianza considerados. Esto es esencial para que el lector se asegure de que el estudio haya obtenido suficiente precisión estadística (144).

7.1.4 Resultados. Ítems 13-17

La sección de resultados debe dar cuenta en profundidad de la tasa de respuesta y de la descripción de la población de estudio, junto con los principales resultados descriptivos y analíticos. La información proporcionada dependerá del tipo de diseño observacional seguido por el investigador y el correspondiente análisis estadístico realizado.

7.1.5 Discusión. Ítems 18-21

La discusión debe abordar todas las cuestiones centrales del estudio, incluida la validez. Los objetivos del estudio deben tenerse en cuenta al discutir los hallazgos.

Las comparaciones con la literatura ya publicada son esenciales. Puede ser apropiado que la sección de discusión se subdivida en diferentes secciones para permitir una mejor interpretación de los hallazgos del estudio. El investigador debe proporcionar una evaluación objetiva de estos y evitar interpretaciones excesivas.

Se deben considerar los posibles efectos de confusión que podrían haber tenido un efecto sobre los resultados y las asociaciones obtenidas en el estudio, por lo que es imperativo anotar las posibles limitaciones que afronta el estudio al tiempo que se observa cualquier sesgo que pudiera haber estado presente. Así mismo, los investigadores deben tener en cuenta que la causalidad de un resultado en particular no se puede establecer en la mayoría de los diseños de estudio, a menos que se haya realizado un estudio de cohorte longitudinal, por lo que este hecho debe reconocerse durante la discusión, pues puede actuar como una limitación del mismo para ciertos diseños de investigación. Las limitaciones del estudio van de la mano con recomendaciones para futuras investigaciones, para validar el estudio o establecer aún más asociaciones presentes en dicho artículo.

7.1.6 Financiación y patrocinio. Ítem 22

La fuente de financiación y el papel de los financiadores en el estudio son piezas de información esenciales que se requieren al final del artículo. Esto va acompañado de cualquier conflicto de intereses, tanto de los autores como de los patrocinadores.

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported
Objectives	3	State specific objectives, including any prespecified hypotheses
Methods		
Study design	4	Present key elements of study design early in the paper
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection
Participants	6	(a) <i>Cohort study</i> —Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up <i>Case-control study</i> —Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls <i>Cross-sectional study</i> —Give the eligibility criteria, and the sources and methods of selection of participants (b) <i>Cohort study</i> —For matched studies, give matching criteria and number of exposed and unexposed <i>Case-control study</i> —For matched studies, give matching criteria and the number of controls per case
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group
Bias	9	Describe any efforts to address potential sources of bias
Study size	10	Explain how the study size was arrived at
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) <i>Cohort study</i> —If applicable, explain how loss to follow-up was addressed <i>Case-control study</i> —If applicable, explain how matching of cases and controls was addressed <i>Cross-sectional study</i> —If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses

Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed (b) Give reasons for non-participation at each stage (c) Consider use of a flow diagram
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders (b) Indicate number of participants with missing data for each variable of interest (c) <i>Cohort study</i> —Summarise follow-up time (eg, average and total amount)
Outcome data	15*	<i>Cohort study</i> —Report numbers of outcome events or summary measures over time <i>Case-control study</i> —Report numbers in each exposure category, or summary measures of exposure <i>Cross-sectional study</i> —Report numbers of outcome events or summary measures
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses
Discussion		
Key results	18	Summarise key results with reference to study objectives
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence
Generalisability	21	Discuss the generalisability (external validity) of the study results
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based

Figura 13. Lista de verificación STROBE.

Poblaciones de estudio

8. POBLACIONES DE ESTUDIO

8.1 DIABETES

8.1.1 Definición, tipos de diabetes y diagnóstico

La diabetes mellitus (DM) fue definida por la Organización Mundial de la Salud (OMS) como “*un trastorno metabólico de etiología múltiple caracterizado por hiperglucemia crónica con alteraciones del metabolismo de carbohidratos, grasas y proteínas como resultado de defectos en la secreción de insulina, acción de la insulina o ambos*” (154).

Existen diversos tipos de diabetes, los 2 más habituales son la *diabetes mellitus tipo 1 (DM1)* y la *diabetes mellitus tipo 2 (DM2)*. La DM1 es un trastorno crónico que se caracteriza por una secreción de insulina insuficiente o total por parte del páncreas (evidencia de destrucción autoinmune de las células β pancreáticas), mientras que la DM2 es el resultado de la combinación de una inadecuada respuesta secretora compensatoria y una resistencia a la acción de la insulina, asociada a la obesidad en la mayoría de los casos (155). Esta última es, además, más común.

En la actualidad se emplean 4 pruebas diagnósticas para la diabetes:

1) Medición de la glucosa plasmática en ayunas; diagnóstico confirmado si valores de glucosa plasmática en ayunas ≥ 7.0 mmol / L (126 mg / dl).

2) Prueba de tolerancia oral a la glucosa (PTOG) (glucosa plasmática poscarga 2 horas después de una prueba de tolerancia oral a la glucosa de 75 g); diagnóstico confirmado si glucosa plasmática 2 h después de la carga ≥ 11.1 mmol / L (200 mg / dl) (5).

3) Hemoglobina glicosilada (HbA1c); diagnóstico confirmado si HbA1c $\geq 6.5\%$ (48 mmol / mol).

4) Glucemia aleatoria en presencia de signos y síntomas de diabetes; diagnóstico confirmado si glucosa en sangre aleatoria ≥ 11.1 mmol / L (200 mg / dl) en presencia de signos y síntomas (156). Si se detectan valores elevados en

personas asintomáticas, se recomienda repetir la prueba (preferiblemente la misma prueba) tan pronto como sea posible para confirmar el diagnóstico (156).

8.1.2 Epidemiología y carga global de la diabetes

La diabetes es una patología presente en todas las regiones y poblaciones del mundo, incluidas las zonas rurales de los países de ingresos medios y bajos. El número de diabéticos está aumentando constantemente y la OMS estima que había 422 millones de adultos con esta patología en todo el mundo en 2014 (157). En 2016, aproximadamente 1.6 millones de muertes fueron causadas directamente por esta enfermedad, 2.2 millones fueron atribuibles a cifras altas de glucemia y casi la mitad de todas ellas se produjeron antes de los 70 años (158). Esas cifras altas de glucosa mantenidas en sangre producen casi 4 millones de muertes cada año (7) y, sin las intervenciones adecuadas, se calcula que al menos 629 millones de personas vivirán con diabetes en el año 2045 (159). En cuanto al gasto económico a nivel mundial en atención médica ocasionado por esta patología, según la Federación Internacional de Documentación (FID), en población adulta fue aproximadamente de 850 mil millones de dólares en 2017 (159).

8.1.3 Etiopatología y consecuencias de la diabetes

Actualmente existe un consenso generalizado basado en que la característica subyacente común a todas las formas de diabetes es la disfunción o destrucción de las células β pancreáticas (160–163). Muchos mecanismos pueden conducir a una disminución de la función o la destrucción completa de dichas células (estas células no se reemplazan, ya que el páncreas humano parece incapaz de renovarlas después de los 30 años de edad) (164). Estos mecanismos incluyen una predisposición genética y anomalías, procesos epigenéticos, resistencia a la insulina, autoinmunidad, enfermedades concurrentes, inflamación, factores ambientales y una asociación entre las variantes genéticas específicas y la respuesta a las terapias hipoglucemiantes (farmacogenómica) (165).

Algunos de los efectos específicos a largo plazo de la diabetes incluyen retinopatía, nefropatía y neuropatía entre otras complicaciones. Las personas con

diabetes también tienen un mayor riesgo de padecer otras patologías, como enfermedades cardíacas, arteriales periféricas y cerebrovasculares, obesidad, cataratas, disfunción eréctil y enfermedad del hígado graso no alcohólico. Puede presentarse con síntomas característicos como sed, poliuria, visión borrosa y pérdida de peso. Sus manifestaciones clínicas más graves son la cetoacidosis o un estado hiperosmolar no cetósico (166).

8.1.4 Tratamiento no farmacológico

- *Recomendaciones generales*

El tratamiento debe tener como objetivo principal disminuir los niveles de glucemia, para así evitar descompensaciones agudas, síndrome hiperosmolar o cetoacidosis, apaciguar los síntomas cardinales (poliuria, polidipsia, polifagia, astenia), minimizar el riesgo de desarrollo o progresión de retinopatías, nefropatías y/o neuropatías (diabéticas), optimizar el perfil lipídico, evitar hipoglucemias y reducir la mortalidad (154).

- *Educación diabetológica*

Un equipo de educación para la diabetes ideal debe estar constituido por un endocrino, una enfermera especializada en diabetes y un dietista, con la eventual colaboración de un psicólogo y una trabajadora social (el equipo mínimo de profesionales para fomentar la educación en el paciente diabético deberá estar integrado por un médico y una enfermera, preferiblemente dentro del ámbito de la asistencia primaria) (167–169). El contenido del programa de educación para la diabetes debe individualizarse en función del tipo de diabetes, la presencia de complicaciones y el nivel sociocultural del paciente (170).

En la mayoría de los casos, alcanzar esos niveles de glucemia recomendados precisa de un programa de adiestramiento por parte del paciente para un buen control de su enfermedad, fundamentado en mediciones frecuentes del nivel de glucemia capilar, recomendaciones nutricionales, actividad física, indicaciones para evitar y solventar las hipoglucemias, un régimen de insulina adecuado a su estilo de vida y evaluaciones periódicas de los resultados adquiridos (171).

- *Recomendaciones nutricionales*

El conocimiento por parte de los pacientes sobre los principios de la nutrición y la elección de los alimentos es un pilar esencial en el manejo de la diabetes, y en la mayoría de ocasiones la única intervención necesaria (172). Se trata de uno de los factores más efectivos en la prevención a corto plazo y el aplazamiento a largo plazo de las complicaciones generadas por esta patología (173–175). Por ello, una dieta adecuada debe tener como objetivo fundamental conseguir un peso apropiado y unos niveles recomendables de glucosa, además de optimizar otros marcadores como los lípidos y la hipertensión arterial (176–178). El tipo de dieta debe ser similar al recomendado para el resto de la población: dieta DASH (Dietary Approaches to Stop Hypertension), dieta mediterránea, dieta saludable nórdica, una restricción moderada de carbohidratos y prescindir de azúcares solubles y derivados por su rápida absorción (172). Es importante repartir las calorías consumidas a lo largo del día en 4 o 5 comidas. El DM1 apenas tendrá que modificar sus hábitos alimenticios (179), mientras que los diabéticos con obesidad (generalmente DM2) requerirán una dieta de tipo hipocalórico (180). En aquellos pacientes en tratamiento con insulina es imprescindible que exista una regularidad y una sincronía en la cantidad y composición de las comidas, los horarios y la farmacocinética del tipo de insulina que se use.

Por último, la dieta también tiene que ser útil en la prevención y el tratamiento de episodios agudos que puedan darse en el paciente diabético (hipoglucemias), así como posibles complicaciones crónicas (172).

- *Actividad física*

En la DM1, el deporte tiene que reportar los mismos beneficios y entretenimiento que para el no diabético, ya que va a servir como un complemento tanto para la dieta como para la acción de la insulina. El sujeto debe tener en cuenta que la actividad física puede ampliar el riesgo de sufrir una hipoglucemia (episodio agudo o diferido), por lo que debe adaptar convenientemente su dieta y la dosis de insulina cuando se disponga a practicarlo o lo haya acabado, considerando duración, intensidad del mismo y cifras de glucemia (181).

En la DM2, la actividad física juega un papel crucial favoreciendo el aumento de la captación de glucosa a través del músculo, inclusive cuando no se reduce el peso, ayudando a optimizar el control del metabolismo. Además, interviene de una manera favorable atenuando otros factores de riesgo de patología cardiovascular como la hipertensión arterial y la hiperlipidemia.

Las recomendaciones se basan en ejercicios de tipo moderado como caminar (50 minutos por sesión de caminata rápida), natación o bicicleta. La intensidad de los mismos deberá incrementarse poco a poco, y se realizará el ejercicio al menos 3 días por semana con el fin de favorecer la sensibilidad a la insulina (182).

8.1.5 Tratamiento farmacológico

- *Fármacos orales*

Una vez fracasada la dieta y la actividad física, los fármacos orales son el tratamiento de elección en casi todos los diabéticos. No existe aún una pauta clara sobre cuál emplear basándonos en las características del paciente, pero sí se han planteado numerosos algoritmos para un adecuado control terapéutico. Habitualmente, la elección del fármaco para el inicio del tratamiento se hace en relación con el nivel de glucemia y el índice de masa corporal (IMC) que presenta el sujeto (183) (figura 14).

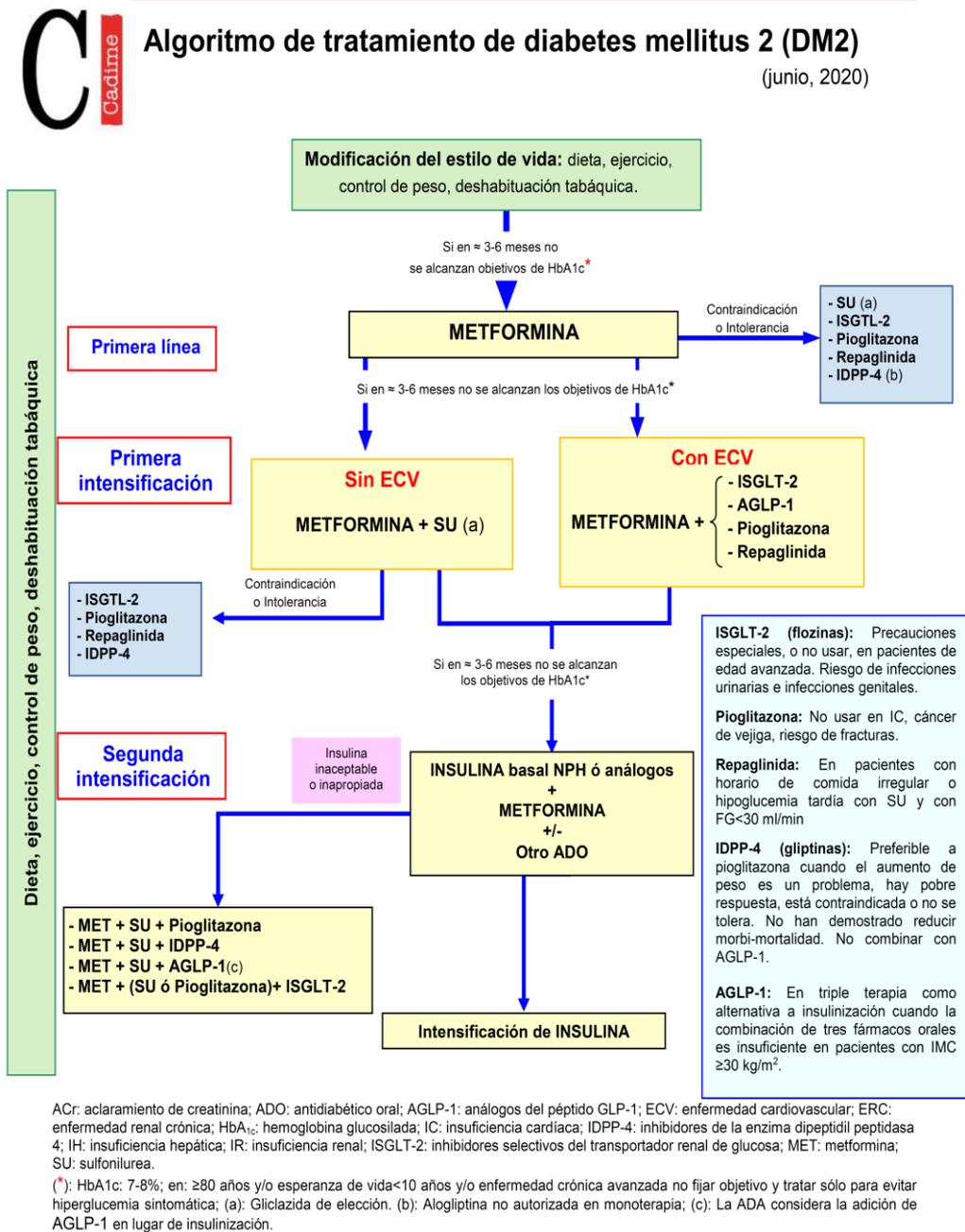


Figura 14. Algoritmo para el tratamiento de la DM2.

- *Insulina*

La insulina se prescribe en los siguientes casos: en el momento del diagnóstico para pacientes que no presentan obesidad, en presencia de síntomas característicos y cifras de glucemia elevadas en pacientes diabéticos con fracaso primario o secundario a la combinación de fármacos vía oral y en situación de descompensación en fase aguda por patologías concomitantes (184).

8.1.6 Autocontrol

En la actualidad, la existencia de dispositivos para medir la glucemia capilar (muestra de sangre venosa (plasma), monitorización continua de glucosa (MCG), dispositivo de glucemia capilar (DGC)) permiten un minucioso autocontrol por parte del paciente. La periodicidad para medir estos niveles de glucemia vendrá determinada por el tipo de diabetes y el tipo de régimen terapéutico para promover la autonomía, y la responsabilidad en el tratamiento y control de su patología (185). Otro método que aún sigue siendo efectivo es el utilizado para determinar la presencia de cuerpos cetónicos en la orina en casos de enfermedad aguda, niveles mantenidos de glucemia > 300 mg/dl, embarazo, estrés o síntomas de cetoacidosis (186,187).

8.1.7 Cambios en el estilo de vida

Existe una alta evidencia para afirmar que la DM2 se puede prevenir modificando el estilo de vida mediante un cambio en la dieta (de acuerdo con las recomendaciones actuales en términos de calidad de grasa, ingesta de fibra, mayor uso de productos integrales, frutas y verduras) y un aumento de la actividad física. La reducción del riesgo de DM2 está fuertemente relacionada con el grado de pérdida de peso a largo plazo y la adherencia a los cambios en el estilo de vida. Se ha demostrado que este efecto preventivo se mantiene durante muchos años después de la intervención activa (188,189).

Para garantizar la sostenibilidad, las modificaciones de comportamiento deben ser adoptadas por los pacientes y respaldadas por sus familiares y amigos. La relación con un grupo de actividad física, sentirse competente y exitoso en la gestión del cambio y el apoyo de los profesionales de la salud parecen ser factores

clave para adherirse a un nuevo estilo de vida (190). Además, coexisten otros factores como la motivación (imprescindible para lograr dichos cambios), la consecución de resultados, el afrontamiento de los desafíos actuales, el apoyo de los profesionales de la salud, la asistencia y afinidad social y la identificación y aceptación del estilo de vida (190).

8.2 CÁNCER

8.2.1 Definición

El término *cáncer*, conocido también por *tumor maligno* o *neoplasia maligna*, hace referencia de forma genérica a un extenso grupo de enfermedades que afectan a cualquier parte del organismo. La característica que lo define es la capacidad de multiplicación de las células anormales y su extensión más allá de los límites habituales, con la consiguiente invasión de las diferentes partes adyacentes del cuerpo o propagación a otros órganos, proceso conocido como *metástasis*, siendo esta su principal causa de muerte (191).

8.2.2 Epidemiología

Se estima que hubo 19.3 millones de nuevos casos de cáncer y casi 10 millones de muertes en todo el mundo en el año 2020. Los cánceres diagnosticados con mayor frecuencia fueron el cáncer de mama femenino (2.26 millones de casos), el cáncer de pulmón (2.21) y el de próstata (1.41), siendo las causas más comunes de muerte el cáncer de pulmón (1.79 millones de muertes), el de hígado (830000) y el de estómago (769000) (192).

8.2.3 Etiología y factores de riesgo

El cáncer es la consecuencia de la mutación de un grupo de células normales en células tumorales, pasando por una serie de fases en las que una lesión precancerosa evoluciona finalmente a un tumor maligno. Esas variaciones resultan de la interacción entre 3 categorías de agentes externos y determinados factores genéticos: carcinógenos físicos (radiaciones ultravioletas y ionizantes),

carcinógenos químicos (amianto, humo del tabaco, aflatoxinas y arsénico) y carcinógenos biológicos (virus, bacterias y parásitos). Esta clasificación de los agentes cancerígenos fue dada por la OMS a través de un órgano especializado, el Centro Internacional de Investigaciones sobre el Cáncer (CIIC) (193).

La edad aumenta considerablemente la incidencia de esta enfermedad, posiblemente porque se van sumando numerosos factores de riesgo relacionados con ciertos tipos de cáncer. A esto hay que añadir también un deterioro de la eficacia de los mecanismos de reparación a nivel celular (194).

Algunos de esos factores de riesgo son el consumo de tabaco y alcohol, la inactividad física, una dieta poco saludable y la contaminación del aire. Determinadas infecciones crónicas son consideradas también factores de riesgo para desarrollar un cáncer (un factor problemático en países de ingresos medios y bajos) (195,196). En el mundo, en el año 2018, aproximadamente un 13% de los casos de cáncer diagnosticados fueron atribuidos a infecciones, principalmente producidas por el *Helicobacter pylori*, los virus de la hepatitis B y C, VPH y el virus de Epstein-Barr (197).

8.2.4 Prevención frente al cáncer

Con el fin de prevenir la aparición del cáncer, debemos disminuir el riesgo evitando el consumo de tabaco y alcohol a través de una dieta saludable rica en frutas y verduras, manteniendo un normopeso, realizando actividad física de forma regular, vacunando a la población de riesgo frente al VPH y la hepatitis B, previniendo la radiación ultravioleta, y reduciendo la exposición a la contaminación del aire exterior e interior y a las radiaciones ionizantes (196).

8.2.5 Detección temprana

Es de vital importancia una detección precoz para potenciar la eficacia del tratamiento y así poder reducir la mortalidad por cáncer. Para la detección temprana debemos de tener en cuenta dos aspectos fundamentales: un diagnóstico precoz y un cribado (198).

- *Diagnóstico precoz*

Un diagnóstico temprano de los cánceres sintomáticos es ventajoso en la mayoría de los casos para prácticamente la totalidad de los tipos de cáncer. Los planes oncológicos deben estar diseñados para disminuir los retrasos y aquellos inconvenientes que interfieran en el uso de los servicios relacionados con el diagnóstico, tratamiento y cuidados necesarios (199).

Cuando existe un diagnóstico precoz se mejora notablemente la vida de los pacientes, la probabilidad de que el tratamiento sea eficaz es mayor, al igual que la probabilidad de supervivencia, de tal modo que la morbilidad se reduce y además se abarata el tratamiento (200).

Para garantizar el éxito de un diagnóstico temprano debemos de tener en cuenta 3 factores: tener conocimientos sobre los síntomas de las variadas formas de cáncer y buscar consejo médico si existen motivos para ello, tener acceso a los servicios de evaluación clínica y diagnóstico, y efectuar la derivación pertinente del paciente a los servicios de tratamiento específicos (191).

- *Cribado*

El objetivo del cribado es captar a aquella población con predisposición a padecer un cáncer o una lesión precancerosa en función de unos resultados que muestren esa condición antes de que aparezcan los síntomas. Cuando se hallan anomalías mientras se realiza el cribado deben efectuarse más pruebas para determinar o descartar un diagnóstico, y poder así aplicar al paciente un tratamiento en caso de que este fuera necesario. Los programas de cribado existentes son efectivos para determinados tipos de cáncer (no para todos), y generalmente son bastante más complejos y caros que el diagnóstico precoz (se necesitan equipos especiales y personal cualificado) (201).

En los programas de cribado se recluta a los pacientes en función de la edad y el riesgo para impedir una realización masiva de pruebas diagnósticas que puedan dar lugar a falsos positivos. Algunos de esos métodos de cribado son la citología vaginal y la prueba del VPH para detectar el cáncer de cuello uterino, la mamografía para el cribado del cáncer de mama, la inspección visual con ácido acético para detectar el cáncer cervicouterino (202–204), o la detección de sangre

oculta en heces, sigmoidoscopia y colonoscopia para el diagnóstico del cáncer colorrectal (205).

8.2.6 Tratamiento

Cada cáncer precisa de un adecuado diagnóstico y una terapia determinada para lograr el tratamiento más apropiado y eficaz. En la actualidad ese tratamiento se basa en quimioterapia, radioterapia y/o cirugía (206–208). El principal objetivo es curar el cáncer o, en su defecto, prolongar la vida en la medida de lo posible basándose en unas consideraciones éticas. Establecer los objetivos del tratamiento es quizá el paso más importante junto con el de mejorar la calidad de vida del paciente, algo factible de alcanzar a través del bienestar físico, psicosocial y espiritual del individuo y mediante cuidados paliativos en las etapas finales de la enfermedad (209–214).

Dentro de los cánceres más recurrentes están el cervicouterino, el de mama, el colorrectal, el bucal, el seminoma, y algunos tipos de leucemias y linfomas en niños. Cuando estos cánceres son detectados a tiempo y reciben el tratamiento adecuado, presentan tasas de curación elevadas (192).

8.2.7 Cuidados paliativos

El propósito de los cuidados paliativos nunca será el de curar el cáncer, sino aliviar los síntomas causados por la enfermedad y mejorar la calidad de vida de los pacientes y de sus familiares. Es imprescindible aplicarlos en aquellos lugares donde existe una elevada tasa de sujetos en estadios avanzados de la enfermedad y reducidas posibilidades de curación. Estos cuidados posibilitan que más del 90% de los pacientes que se encuentran en una de esas etapas finales encuentren alivio a sus problemas físicos, psicosociales y espirituales (209–214). Existen una serie de estrategias de salud pública que facilitan la atención a enfermos y familiares no solo dentro del ámbito hospitalario, sino también en su propio domicilio, esenciales para brindar unos cuidados paliativos individualizados y personalizados (215,216).

Es recomendable favorecer el acceso del paciente a ciertos tipos de fármacos de administración vía oral, sobre todo la morfina, para mitigar esos dolores

moderados o intensos causados por la enfermedad que se presentan en más del 80% de estos pacientes en fase terminal (217).

8.2.8 Eutanasia

En España, aquellas personas que sufran una enfermedad grave e incurable ya pueden solicitar la eutanasia y, por lo tanto, podrán decidir entre recibir cuidados paliativos para mejorar su calidad de vida al final de esta o solicitar una muerte asistida. El tipo de eutanasia aprobado en España es el activo directo, que se produce cuando “la muerte es consecuencia directa de la acción de un tercero” (suicidio médicamente asistido) (218). El paciente tendrá que ser informado de las diferentes alternativas disponibles a esta práctica y tendrá la obligación de ratificar su voluntad de morir al menos 4 veces. La eutanasia se puede realizar en centros de salud o en el domicilio del paciente (219).

La ley en cuestión que permite la legalización de la eutanasia activa directa en España y su inserción en la cartera común de servicios del Sistema Nacional de Salud, es la Ley Orgánica 3/2021, de 24 de marzo, de regulación de la eutanasia (220):

“Podrá solicitar la prestación de ayuda incluida en el Sistema Nacional de Salud de la eutanasia cualquier persona mayor de edad y en plena capacidad de obrar y decidir, siempre que lo haga de forma autónoma, consciente e informada. Debe tener la nacionalidad española, residencia legal en España o certificado de empadronamiento que acredite un tiempo de permanencia en territorio español superior a doce meses” (221).

8.3 DISFUNCIÓN DEL PROCESAMIENTO SENSORIAL

8.3.1 Definición

Para Jane Ayres el proceso de integración sensorial se define como un “proceso neurológico que organiza las sensaciones del propio cuerpo y del medio ambiente y hace posible usar el cuerpo efectivamente en el entorno” (222). Se produce a nivel cerebral y posibilita organizar, analizar e interpretar las diferentes sensaciones que recibimos a través de los sentidos para ejercer y desarrollar nuestras

actividades de la vida diaria. Dicho proceso es bastante complejo y supone la ejecución de cuantiosos mecanismos por parte del sistema nervioso central (223,224).

8.3.2 Etiología y consecuencias

El modo de procesar la información de cada persona depende de diversos factores ambientales, biológicos, genéticos y experiencias vitales, produciéndose entre ellos una interacción dinámica que explica el origen multicausal de las disfunciones sensoriales. Cuando existe un procesamiento sensorial limitado, se producen una serie de conductas disfuncionales, como problemas en la regulación del estado de atención, alerta y/o sueño; dificultades en la ejecución de actividades y ocupaciones; retraso en el desarrollo de habilidades; trastornos en el autoconocimiento y la autoestima y en la conducta o las emociones (225).

8.3.3 Disfunciones de la integración sensorial

Se producen cuando el procesamiento sensorial (natural, automático e inconsciente) se ejecuta de una forma deficiente, es decir, cuando el sistema nervioso central es incapaz de procesar, organizar e integrar la información sensorial de forma adecuada. Existe una gran variabilidad, y esta se puede presentar de manera conjunta o combinada.

Las disfunciones pueden darse con relación a una o varias de sus etapas y se clasifican en 3 categorías (226):

- Trastorno de la Modulación Sensorial (TMS): cuando el problema se produce por un registro sensorial limitado o nulo y/o una modulación sensorial incorrecta. Las alteraciones tienen lugar para los niveles de alerta y actividad. Existen 3 subtipos: hiporrespuesta, hiporresponsividad o hiposensibilidad sensorial (no se registra de manera adecuada el estímulo sensorial); hiperrespuesta, hiperresponsividad o hipersensibilidad al input sensorial (problemas en la modulación sensorial), e interés o búsqueda sensorial (226–231).
- Trastorno sensitivo-motor (TSM): cuando se muestra una respuesta motora desorganizada como resultado del procesamiento incorrecto de la

información sensorial que afecta al control postural, lo que deriva en un trastorno postural o del desarrollo de la coordinación. Existen 2 subtipos: dispraxia (el problema está relacionado con la discriminación y/o con la combinación de un estímulo con otros, y las alteraciones se producen en el ámbito de la planificación motora, secuenciación y organización temporo-espacial) y trastorno postural (226,232–234).

- Trastorno de discriminación sensorial (TDM), supone un procesamiento erróneo de la información sensorial (226). Existen 7 subtipos: visual, auditivo, táctil, gustativo, olfativo, vestibular y propioceptivo (235). J. Ayres estudió cómo influyen los sistemas sensoriales vestibular, propioceptivo y táctil en el desempeño ocupacional, y detalló las consiguientes disfunciones con respecto a estos 3 sentidos (222,236,237).

- *Disfunción vestibular*

El cerebro no es capaz de procesar de manera correcta la información sensorial vestibular, reaccionando de una forma desmedida o, por el contrario, de una forma leve (237). Encontramos 2 subtipos: hiposensibilidad al estímulo vestibular, caracterizada por un pobre registro del input vestibular (dificultades para el control postural, actividades bilaterales y secuenciación. En niños, problemas de atención, dificultades académicas y de motricidad fina), e hipersensibilidad al estímulo vestibular, en la que el niño reconoce el estímulo, pero no es capaz de modular su intensidad. Este a su vez presenta 2 subtipos: la intolerancia o aversión al movimiento y la inseguridad gravitacional (238–240).

- *Disfunción propioceptiva*

El cerebro no es capaz de procesar de un modo adecuado la información que percibe de su propio cuerpo a través de los músculos y las articulaciones. Podemos encontrar 2 subtipos: hiposensibilidad al estímulo propioceptivo (el cerebro registra de forma insuficiente o no registra los inputs propioceptivos, induciendo en el sujeto dificultades motoras, tono muscular bajo, torpeza, disociación, dificultades de coordinación (niños que se muestran excesivamente “blandos”, como si les “pesara” su propio cuerpo)) y propiocepción como modulador (niños habitualmente denominados “buscadores”, exploran input

propioceptivos para compensar otros sistemas sensoriales a los que son hipersensibles con el fin de tranquilizarse (niños tachados de “agresivos”, “hiperactivos” o “problemáticos” porque presentan conductas dañinas para sí mismos o para los demás)) (225,239,241–244).

- *Disfunción táctil*

Afectación que produce alteraciones de integración en el sistema táctil. Podemos diferenciar 3 subtipos: hiporresponsividad al estímulo táctil, relacionada con los problemas para almacenar la información táctil (niños que buscan acariciar, tocar, abrazar, entrar en contacto); defensividad táctil, consiste en estímulos táctiles supuestamente inocuos que son registrados con enorme intensidad, resultando molestos y desagradables (dificultades importantes para la realización de actividades de la vida diaria, como el vestido, la ducha o la comida), y trastornos de discriminación táctil, caracterizados por una dificultad para la identificación, organización y empleo de los inputs táctiles (problemas para realizar actividades que incluyan motricidad fina y coordinación, conocimiento insuficiente de su propio cuerpo, torpeza motora o problemas de praxis) (225,242,245,246).

- *Dispraxia*

Trastorno psicomotriz caracterizado por la dificultad en los movimientos que precisan la movilización y coordinación de varios grupos musculares y su capacidad para resolver nuevas situaciones (niños que poseen altas habilidades cognitivas, pero que tienen dificultades para resolver actividades especializadas nuevas, suelen ser etiquetados de “torpes” o “vagos” porque a los profesores o padres les resulta difícil concebir que un niño con recursos cognitivos se demore o no ejecute bien tareas fáciles de realizar) (242,247–249).

8.3.4 Evaluación en integración sensorial

El Perfil Sensorial posibilita evaluar los problemas sensoriales relacionados con los sujetos desde su nacimiento hasta su senectud. Encontramos 5 perfiles diferentes (223):

- Adolescent/Adult Sensory Profile
- Sensory Profile
- Sensory Profile Supplement
- Sensory Profile School Companion
- Infant/Toddler Sensory Profile

Dichos perfiles son aplicados en función del rango de edad de la población de referencia, aportando un mayor conocimiento sobre las necesidades del procesamiento sensorial y pudiendo identificar estas dentro de las situaciones cotidianas.

El Adolescent/Adult Sensory Profile (AASP) (250), cuestionario empleado en el artículo de investigación de nuestra tesis doctoral titulado "Validation of the Adolescent/Adult Sensory Profile (AASP) for Spain", es un cuestionario autoadministrado que consta de 60 ítems que evalúan los niveles de procesamiento sensorial para la vida diaria a través de 6 secciones (gusto / olfato, movimiento, visual, táctil, actividad y auditiva). La totalidad de los ítems se centran en tres áreas principales: la primera área valora la capacidad de regular las entradas sensoriales que afectan a las actividades de la vida diaria (AVD). La segunda evalúa la capacidad de interpretar distintas entradas sensoriales relacionadas con diferentes sistemas sensoriales. Finalmente, la tercera evalúa la capacidad para procesar los estímulos sensoriales y responder ante ellos (251). Se debe indicar la frecuencia de las respuestas a las distintas experiencias sensoriales para una escala de 5 puntos; 1 (casi nunca), 2 (raras veces), 3 (ocasionalmente), 4 (con frecuencia), y 5 (casi siempre). El cuestionario tarda en completarse de 10 a 15 minutos. Está basado en el modelo de procesamiento sensorial de Dunn (1997) (252), los principios interactivos de los umbrales neurológicos (pueden ser altos o bajos) y las respuestas de tipo conductual (pueden ser activas o pasivas). AASP divide los resultados en 4 probables cuadrantes en relación con los perfiles de comportamiento: "registro bajo", "búsqueda sensorial", "sensibilidad sensorial" y "evitación sensorial".

Cuando se compara entre la integración sensorial de Ayres y la teoría de procesamiento sensorial de Dunn, debe haber un acercamiento hacia lo que se

conoce como dominio de Terapia Ocupacional dentro del Marco de Trabajo de Terapia Ocupacional, basado en Destrezas de Ejecución, Demandas de la Actividad, Características del Cliente, Patrones de Ejecución, Áreas de Ocupación y Contexto, y Entorno (253).

Desde la Teoría de Integración Sensorial se lleva a cabo un acercamiento a las Destrezas de Ejecución y las Características del Cliente (tanto la evaluación terapéutica como el proceso de intervención se orientan a las destrezas motoras y perceptuales, funciones del cuerpo, destrezas cognitivas y de regulación emocional), mientras que cuando se examina la teoría del Procesamiento Sensorial se dirige esencialmente a los Contextos y Entornos y a las Demandas de la Actividad (importante consideración de las demandas del espacio y tiempo en la ocupación, de las expectativas sociales, del empleo del cuerpo y los objetos en relación con estos, y de las demandas del entorno y el contexto que pueden alterar el desempeño a la hora de participar en las actividades) (253).

8.3.5 Intervención en integración sensorial

El modelo de intervención establecido por la Dra. Ayres se fundamenta en los 3 sistemas sensoriales anteriormente citados: vestibular, propioceptivo y táctil (teniendo en cuenta también los sistemas visual y auditivo) (223).

La intervención va a venir dada por el tipo de disfunción en el procesamiento sensorial, además de las características individuales de cada sujeto, como la edad, el entorno, las dificultades presentes para su desempeño ocupacional y sus fortalezas (254).

Es un requisito imprescindible y previo al tratamiento el realizar una evaluación individualizada y exhaustiva. Si posterior a dicha evaluación el terapeuta ocupacional establece que el sujeto presenta algún desorden en el procesamiento sensorial, deberá planificar la intervención más apropiada.

Para la Dra. Ayres *“no existe una receta para una sesión de terapia perfecta, y esta varía de un sitio a otro y de una sesión a otra”* (222). Para ello, pormenorizó una sucesión de principios esenciales que debían acontecer en toda intervención y sesión de Integración Sensorial, estos quedaron ampliamente catalogados en la Medida de Fidelidad ASI (Ayres Sensory Integration) (244,246,255–257).

IV. OBJETIVOS

IV. OBJETIVOS

OBJETIVO GENERAL:

- Conocer y aplicar de una forma precisa la metodología COSMIN para la realización de revisiones sistemáticas sobre herramientas de evaluación y para la validación de cuestionarios.

OBJETIVOS ESPECÍFICOS:

- Llevar a cabo una revisión sistemática siguiendo la metodología COSMIN sobre herramientas de evaluación para medir el estado nutricional en pacientes con diabetes mellitus tipo 1 y/o 2, sus propiedades psicométricas y la calidad metodológica de los estudios de validación de estos instrumentos.
- Realizar una revisión sistemática siguiendo la metodología COSMIN sobre herramientas de evaluación para medir el conocimiento dietético en pacientes con diabetes mellitus tipo 1 y/o 2, sus propiedades psicométricas y la calidad metodológica de los estudios de validación de estos instrumentos.
- Efectuar una revisión sistemática siguiendo la metodología COSMIN sobre herramientas de evaluación para medir la fatiga relacionada con el cáncer en cuidados paliativos, sus propiedades psicométricas y la calidad metodológica de los estudios de validación de estos instrumentos.
- Validar el cuestionario Adolescent/Adult Sensory Profile (AASP) y estudiar sus propiedades psicométricas en población española.

V. MARCO EMPÍRICO

Artículo 1

Nutritional status measurement instruments: a systematic psychometric review

*PUBLICADO EN INTERNATIONAL JOURNAL OF ENVIRONMENTAL RESEARCH AND
PUBLIC HEALTH: 7 DE AGOSTO DE 2020.*

Int. J. Environ. Res. Public Health 2020, 17(16), 5719; <https://doi.org/10.3390/ijerph17165719>

NUTRITIONAL STATUS MEASUREMENT INSTRUMENTS: A SYSTEMATIC PSYCHOMETRIC REVIEW

RESUMEN

Antecedentes: la diabetes es una enfermedad crónica grave asociada con un gran número de complicaciones y un mayor riesgo de muerte prematura. Una evaluación dietética es de suma importancia para la promoción de la salud, la prevención de enfermedades y los planes de tratamiento individuales en pacientes con diabetes.

Métodos: Se realizó una búsqueda exhaustiva en diversas bases de datos – Medline, Web of Science, Open Gray, Cochrane Library y COSMIN – para la revisión sistemática de las propiedades de medición de instrumentos que evalúan la ingesta dietética de personas con diabetes mellitus tipo 1 y/o 2 de acuerdo con los estándares COSMIN.

Resultados: Fueron identificados 7 instrumentos. No se contó con un instrumento de medición del estado nutricional para el cual se evaluaran todas las propiedades psicométricas. La calidad metodológica para cada una de las propiedades psicométricas evaluadas fue “inadecuada” o “dudosa” para todos los instrumentos. Food Frequency Questionnaire (FFQ) fue el cuestionario en el que más características psicométricas se evaluaron y con mejor puntuación en cuanto a la calidad de la evidencia.

Conclusiones: Han sido desarrollados varios instrumentos para la evaluación de la ingesta dietética en personas con diabetes. La evaluación de este constructo es de gran utilidad tanto en la práctica clínica como en la investigación, requiriendo nuevos conocimientos en esta área. FFQ es el mejor instrumento disponible para evaluar la ingesta dietética en personas con diabetes.

Palabras clave: estado nutricional; diabetes mellitus; cuestionario.

ABSTRACT

Background: Diabetes is a serious chronic disease associated with a large number of complications and an increased risk of premature death. A dietary evaluation is of utmost importance for health promotion, disease prevention and individual treatment plans in patients with diabetes.

Methods: An exhaustive search was carried out in various databases – Medline, Web of Science, Open Gray, Cochrane Library and COSMIN – for systematic review of the measurement properties of instruments that evaluate the dietary intake of people with diabetes mellitus type 1 and/or 2 according to COSMIN standards.

Results: 7 instruments were identified. There was no instrument measuring nutritional status for which all the psychometric properties were evaluated. The methodological quality for each of the psychometric properties evaluated was “inadequate” or “doubtful” for all instruments. The Food Frequency Questionnaire (FFQ) evaluated the most psychometric characteristics and with a better score in terms of quality of the evidence.

Conclusions: Several instruments have been developed for the evaluation of dietary intake in people with diabetes. Evaluation of this construct is very useful, both in clinical practice and in research, requiring new knowledge in this area. The FFQ is the best instrument available to assess dietary intake in people with diabetes.

Keywords: nutritional status; diabetes mellitus; questionnaire.

1. INTRODUCCIÓN

La diabetes es una enfermedad crónica grave asociada con un gran número de complicaciones y un mayor riesgo de muerte prematura (1). Debido a estas complicaciones, casi la mitad de las muertes (46,1%) correspondieron a adultos menores de 60 años, grupo en edad laboral (2).

La diabetes es un problema de salud pública desafiante, con graves consecuencias para la salud y costes sanitarios asociados a un estilo de vida poco saludable (pérdida de la visión, amputaciones, insuficiencia renal y enfermedades cardiovasculares) (3,4). El control de la diabetes requiere intervenciones farmacológicas y cambios en el estilo de vida que se asocian con un menor riesgo de muerte (5-7).

Mantener una dieta adecuada influye positivamente en la evolución de la diabetes y la aparición de complicaciones (8,9). Los pacientes con diabetes suelen tener dificultades para identificar una dieta adecuada, tanto en calidad como en cantidad (10). La selección de alimentos y el patrón dietético están influenciados por el conocimiento del paciente sobre una dieta recomendada (11). Por lo tanto, la evaluación dietética es de suma importancia para la promoción de la salud, la prevención de enfermedades y los planes de tratamiento individuales en pacientes con diabetes (10,12).

Por esta razón existe la necesidad de contar con herramientas breves que permitan a los profesionales de la salud con conocimientos generales de nutrición evaluar la ingesta dietética de sus pacientes y establecer metas dietéticas basadas en la evaluación obtenida (13,14). En este sentido, existe una amplia variedad de instrumentos validados que pueden utilizarse para una evaluación rápida y precisa (15,16). Estas herramientas no tienen la precisión de una evaluación detallada realizada por un profesional, pero pueden usarse para promover cambios en la dieta cuando se manejan afecciones crónicas (17).

Actualmente, no existen revisiones sistemáticas en la literatura al respecto en relación con esta gran variedad de instrumentos (18) en las que se haya llevado a cabo una evaluación de las propiedades psicométricas y calidad metodológica de los instrumentos basados en COSMIN (COnsensus-based Standards for the selection of health status Measurement INstruments) (19,20). Teniendo en cuenta este hecho, se realizó una revisión sistemática para identificar los instrumentos

para evaluar el estado nutricional en pacientes con diabetes mellitus (DM) tipo 1 y/o 2 (DM1/DM2), valorar sus propiedades psicométricas y evaluar la calidad metodológica de los estudios de validación de estos instrumentos. Los objetivos de esta revisión sistemática fueron: identificar instrumentos para medir el estado nutricional en pacientes con DM1 y/o DM2 y sus propiedades psicométricas y evaluar la calidad metodológica de los estudios de validación de estos instrumentos.

2. METODOLOGÍA

2.1 Diseño

Se efectuó una revisión sistemática de las propiedades de medición de los instrumentos que evalúan la ingesta dietética en personas con DM1 y/o DM2 según la metodología COSMIN para revisiones sistemáticas de los estándares de Medidas de Resultados Informados por el Paciente (PROM) (publicado en febrero de 2018) (19,20).

2.2 Bases de datos y estrategia de búsqueda

2 revisores independientes realizaron una búsqueda exhaustiva en varias bases de datos: Medline (a través de Pubmed), Web of Science, Open Gray, Cochrane Library y COSMIN durante febrero de 2020.

Con el fin de localizar el máximo número posible de herramientas y sus diferentes versiones, no se fijaron fechas. Para realizar esta búsqueda se combinaron los términos MeSH con sinónimos y términos similares (tabla 1), adaptando dicha estrategia a cada uno de los sistemas de búsqueda de las diferentes bases de datos. Además, se revisaron las listas de referencias de los artículos incluidos para identificar otros documentos relevantes. La estrategia de búsqueda se ejecutó siguiendo la metodología COSMIN para revisiones sistemáticas de Medidas de Resultado Informadas por el Paciente (PROMs).

Tabla 1. Estrategia de búsqueda.

1	("Diabetes mellitus" [MeSH] OR "diabetes" [All Fields])
2	("nutritional status"[MeSH] OR "nutrition"[All Fields] OR "nutrition status"[All Fields] OR "eating" [MeSH] OR "food intake" [All Fields] OR "food frequency" [All Fields])
3	("Instrument"[tiab] OR "instruments"[tiab] OR "measure" [tiab] OR "measures" [tiab] OR "questionnaire"[tiab] OR "questionnaires"[tiab] OR "scale"[tiab] OR "scales"[tiab] OR "tool"[tiab] OR "tools"[tiab] OR "survey" [tiab] OR "test" [tiab])
4	("Instrumentation"[sh] OR "methods"[sh] OR "Validation Studies"[pt] OR "Comparative Study"[pt] OR "psychometrics"[MeSH] OR psychometr*[tiab] OR clinimetr*[tw] OR clinometr*[tw] OR "outcome assessment (health care)"[MeSH] OR "outcome assessment"[tiab] OR "outcome measure*[tw] OR "observer variation"[MeSH] OR "observer variation"[tiab] OR "Health Status Indicators"[Mesh] OR "reproducibility of results"[MeSH] OR "reproducib*[tiab] OR "discriminant analysis"[MeSH] OR "reliab*[tiab] OR "unreliab*[tiab] OR "valid*[tiab] OR "coefficient"[tiab] OR "homogeneity"[tiab] OR "homogeneous"[tiab] OR "internal consistency"[tiab] OR ("cronbach*[tiab] AND ("alpha"[tiab] OR "alphas"[tiab])) OR (item[tiab] AND (correlation*[tiab] OR selection*[tiab] OR reduction*[tiab])) OR "agreement"[tiab] OR "precision"[tiab] OR "imprecision"[tiab] OR "precise values"[tiab] OR "test-retest"[tiab] OR ("test"[tiab] AND "retest"[tiab]) OR ("reliab*[tiab] AND ("test"[tiab] OR "retest"[tiab])) OR "stability"[tiab] OR "interrater"[tiab] OR "inter-rater"[tiab] OR "intrarater"[tiab] OR "intra-rater"[tiab] OR "intertester"[tiab] OR "inter-tester"[tiab] OR "intratester"[tiab] OR "intra-tester"[tiab] OR "interobserver"[tiab] OR "inter-observer"[tiab] OR "intraobserver"[tiab] OR "intraobserver"[tiab] OR "intertechician"[tiab] OR "inter-technician"[tiab] OR "intratechnician"[tiab] OR "intra-technician"[tiab] OR "interexaminer"[tiab] OR "inter-examiner"[tiab] OR "intraexaminer"[tiab] OR "intra-examiner"[tiab] OR "interassay"[tiab] OR "inter-assay"[tiab] OR "intraassay"[tiab] OR "intra-assay"[tiab] OR "interindividual"[tiab] OR "inter-individual"[tiab] OR "intraindividual"[tiab] OR "intra-individual"[tiab] OR "interparticipant"[tiab] OR "inter-participant"[tiab] OR "intraparticipant"[tiab] OR "intra-participant"[tiab] OR "kappa"[tiab] OR "kappa's"[tiab] OR "kappas"[tiab] OR "repeatab*[tiab] OR ("replicab*[tiab] OR "repeated"[tiab]) AND ("measure"[tiab] OR "measures"[tiab] OR "findings"[tiab] OR "result"[tiab] OR "results"[tiab] OR "test"[tiab] OR "tests"[tiab])) OR "generaliza*[tiab] OR "generalisa*[tiab] OR "concordance"[tiab] OR ("intraclass"[tiab] AND "correlation*[tiab] OR "discriminative"[tiab] OR "known group"[tiab] OR "factor analysis"[tiab] OR "factor analyses"[tiab] OR "dimension*[tiab] OR "subscale*[tiab] OR "multitrait"[tiab] AND "scaling"[tiab] AND ("analysis"[tiab] OR "analyses"[tiab])) OR "item discriminant"[tiab] OR "interscale correlation*[tiab] OR error[tiab] OR errors[tiab] OR "individual variability"[tiab] OR ("variability"[tiab] AND ("analysis"[tiab] OR "values"[tiab])) OR ("uncertainty"[tiab] AND ("measurement"[tiab] OR "measuring"[tiab])) OR "standard error of measurement"[tiab] OR "sensitiv*[tiab] OR "responsive*[tiab] OR (("minimal"[tiab] OR "minimally"[tiab] OR "clinical"[tiab] OR "clinically"[tiab] AND ("important"[tiab] OR "significant"[tiab] OR "detectable"[tiab]) AND ("change"[tiab] OR "difference"[tiab])) OR ("small*[tiab] AND ("real"[tiab] OR "detectable"[tiab]) AND ("change"[tiab] OR "difference"[tiab])) OR "meaningful change" [tiab] OR "ceiling effect"[tiab] OR "floor effect"[tiab] OR "Item response model"[tiab] OR "IRT"[tiab] OR "Rasch"[tiab] OR "Differential item functioning"[tiab] OR DIF[tiab] OR "computer adaptive testing"[tiab] OR "item bank"[tiab] OR "cross-cultural equivalence"[tiab])
5	#1 AND #2 AND #3 AND #4
6	"Protocol"[Publication Type] OR "addresses"[Publication Type] OR "biography"[Publication Type] OR "case reports"[Publication Type] OR "comment"[Publication Type] OR "editorial"[Publication Type] OR "congresses"[Publication Type] OR "consensus development conference"[Publication Type] OR "consensus development conference"[Publication Type] OR "practice guideline"[Publication Type] OR "suffering from"[tiab] OR "animals"[MeSH]
7	#5 NOT #6
8	FILTER: Language (English and Spanish)
9	FILTER: Species (Humans)

2.3 Criterios de inclusión/exclusión

Los criterios de inclusión fueron: artículos publicados sin límite de tiempo que incluyeran un instrumento que evaluara el estado nutricional en pacientes con DM1 y/o DM2, evaluación de al menos una propiedad de medida de las reflejadas en la lista de verificación COSMIN (validez de contenido, validez estructural, consistencia interna, fiabilidad, error de medición, prueba de hipótesis para validez de constructo, validez transcultural/invariancia de medición, validez de criterio y sensibilidad al cambio) y artículos escritos en inglés o español. Se excluyeron editoriales, biografías, discursos, comentarios y conferencias, informes de casos, guías de práctica, tesis doctorales, artículos de opinión o publicaciones en las que no se utilizó metodología selectiva.

Los títulos y los resúmenes fueron seleccionados de forma independiente por 2 revisores. Se recuperó el artículo completo cuando no se pudieron tomar decisiones sobre el resumen y el título por sí solos. Las discrepancias se resolvieron mediante discusión con un tercer revisor.

2.4 Evaluación de la calidad metodológica

Los estudios incluidos fueron evaluados por Pedro Montagut Martínez (P.M.M.), José Joaquín García Arenas (J.J.G.A.) y David Pérez Cruzado (D.P.C.) empleando la lista de verificación de cuatro puntos COSMIN para revisiones sistemáticas (20).

La lista de verificación COSMIN presenta 9 cajas con 5 - 18 elementos. Cada caja proporciona una puntuación de calidad metodológica para validez de contenido (propiedad que evalúa si cada uno de los elementos incluidos es relevante y completo), validez estructural (mide el grado en que las puntuaciones de un instrumento muestran adecuadamente la dimensionalidad del constructo a medir), consistencia interna (grado de correlación entre los diferentes elementos), validez transcultural (investigación de un fenómeno o teoría que se contrasta en al menos dos culturas diferentes), error de medición (diferencia entre la puntuación empírica de un sujeto y su puntuación real), fiabilidad (estabilidad de medición), validez de criterio (grado en que las puntuaciones de un instrumento corresponden a un "gold standard"), prueba de hipótesis (propiedad que se

analiza para comprobar si las hipótesis se formularon *a priori*) y sensibilidad al cambio (capacidad de una medida para mostrar cambios a lo largo del tiempo).

El contenido evaluado en cada casilla se puede calificar como “muy bueno”, “adecuado”, “dudoso” e “inadecuado”. La puntuación de cada propiedad psicométrica se obtuvo tomando la puntuación más baja de cualquier elemento para cada propiedad (20).

2.5 Extracción de datos

3 autores evaluaron de forma independiente la calidad metodológica de los estudios y las discrepancias se resolvieron por consenso. Todos los artículos fueron seleccionados por revisión de título, resumen y texto completo. De cada artículo se extrajo información sobre los siguientes aspectos: nombre de la herramienta utilizada, población de estudio, ámbito, descripción de la herramienta, propiedades psicométricas evaluadas, puntuación COSMIN y valores estadísticos de las propiedades psicométricas evaluadas.

2.6 Síntesis de datos

Los datos sobre las propiedades psicométricas de los artículos que cumplieron con los criterios de inclusión se calificaron de acuerdo con la lista de verificación COSMIN (20). Los resultados de cada estudio para cada una de las propiedades de medición se calificaron de acuerdo con los criterios de Prinsen (21), clasificando cada resultado como 'positivo' (+), 'negativo' (-) o 'indeterminado' (?). Finalmente, la calidad de la evidencia y la fuerza de las recomendaciones se catalogaron como "alta", "moderada", "baja" o "muy baja", usando el método GRADE (22). Las puntuaciones de cada una de las propiedades psicométricas de cada artículo fueron atribuidas por 3 revisores independientes, y las diferencias se resolvieron mediante discusión y consenso.

3. RESULTADOS

Se identificaron un total de 3538 artículos, de los cuales se examinaron 2217 títulos y resúmenes. De estos, 2195 artículos fueron excluidos tras esta evaluación. Posteriormente, se evaluaron 22 artículos a texto completo y se excluyeron 8 tras

revisar el texto completo (11,23–29). Por lo tanto, se incluyeron un total de 14 estudios en esta revisión sistemática. El diagrama de flujo se muestra en la figura 15.

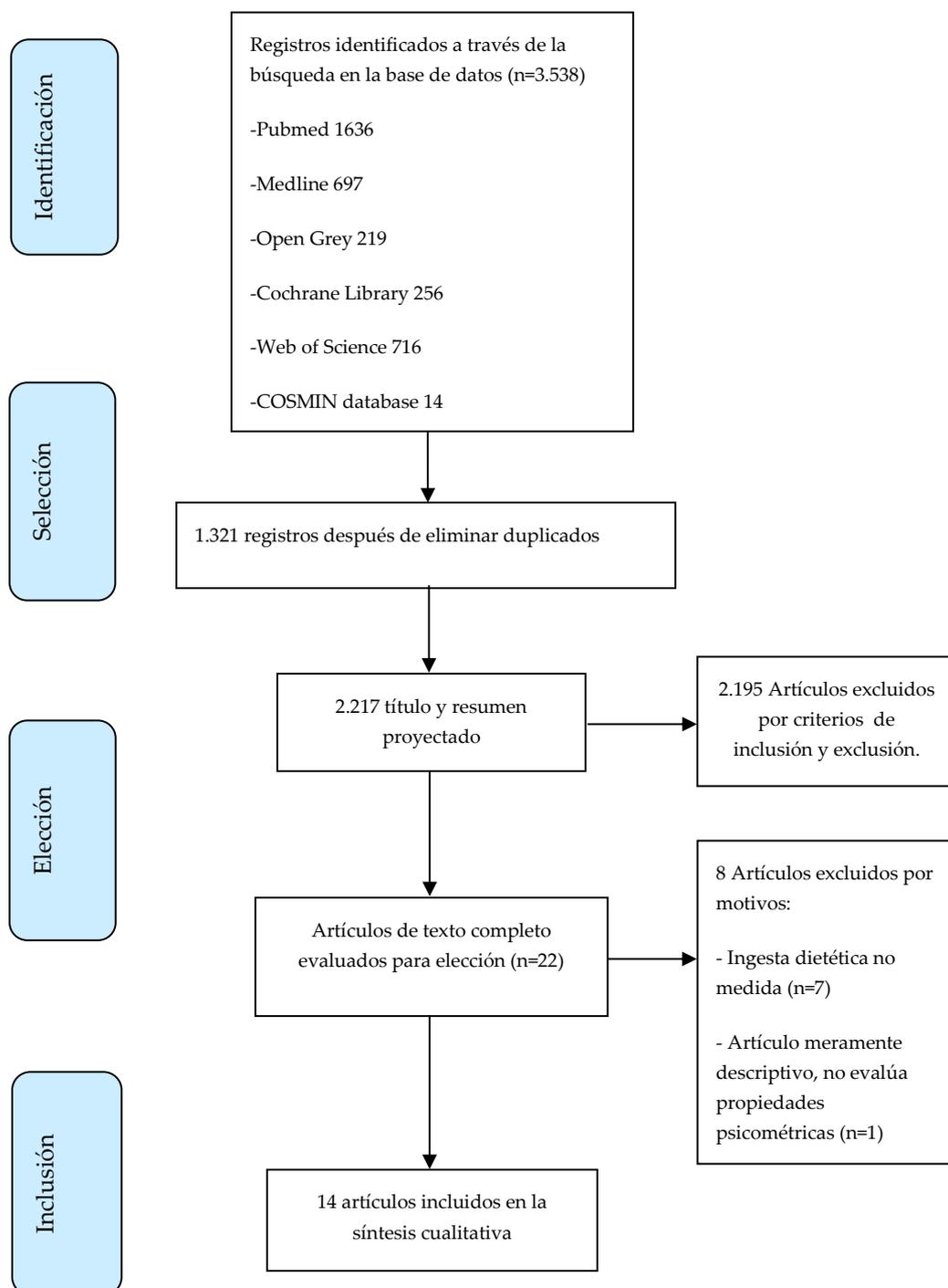


Figura 15. Diagrama de flujo de la selección del estudio.

Se identificaron 7 instrumentos diseñados específicamente para evaluar la ingesta dietética en pacientes con DM1 y/o DM2: Diabetes Knowledge and Behaviour Questionnaire (DKB) (30), Food Frequency Questionnaire (FFQ) (31–40), Perceived Dietary Adherence Questionnaire (PDAQ) (41), UK Diabetes and Diet Questionnaire (UKDDQ) (42), The self-developed Dietary Knowledge Questionnaire (DKQ) (40), Diabetes Mellitus Knowledge Questionnaire (DMK) (40) y Motiv.Diaf-DM2 Questionnaire (MDDM2) (43). En la mayoría de los casos el idioma original utilizado fue el inglés, aunque podemos encontrar varios de estos cuestionarios adaptados a otras versiones: versiones brasileña (32), malí (33), austriaca (34), coreana (35) y china (31) de FFQ, una versión árabe de FFQ, DKQ y DMK (40) y una versión española del Motiv. Diaf-DM2 (43). FFQ varió en el número de elementos y se comparó con un "gold standard" (Weighed Diet Record) en la mayoría de los estudios en los que se empleó. La mayor parte de estas investigaciones se llevaron a cabo en el ámbito hospitalario, comunitario y de atención primaria en pacientes con DM tipo 1 y/o 2 (tabla 2).

Tabla 2. Características y descripción de los instrumentos de medición del estado nutricional en DM.

Estudio (autor y año)	Poplación / tipo de DM	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Diabetes Knowledge and Behaviour Questionnaire (DKB)						
Simmons et al. (1994)	397 adultos con diabetes mellitus tipo 2 (127 europeos, 103 maorís y 167 personas de las islas del Pacífico)	Nueva Zelanda: Los pacientes fueron reclutados de la comunidad.	DKB incluye cinco puntos y 47 preguntas cerradas. Las preguntas se agrupan en 10 apartados con 3-6 respuestas verdaderas / falsas. Para su aplicación definitiva, a la respuesta incorrecta se le asigna un puntaje negativo de -1, la respuesta correcta es un puntaje de +1 y "no sabe" se puntúa con 0. El puntaje total se convierte en un porcentaje con todos los totales negativos calificados como 0 %. El cuestionario lo realiza el entrevistador y su administración demora aproximadamente 30 minutos.	1-Consistencia interna 2- Prueba de hipótesis 3-Sensibilidad al cambio	1-Inadecuado 2-Dudoso 3-Dudoso	- α Cronbach entre 0.59 - 0.90 - El cuestionario de frecuencia alimentaria de 12 ítems basado en una porción estándar se correlacionó bien con la evaluación dietética de las calorías totales ($r = 0.48$ 0.64) y de las calorías debidas a la grasa ($r = 0.41$ -0.65), mientras que el consumo se correlacionó negativamente con el índice de grasa en los europeos ($r = 0.25$, $P < 0.05$) y maorí ($r = -0.33$, $P < 0.01$).

Estudio (autor y año)	Poplación / tipo de DM	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Food Frequency Questionnaire (FFQ)						
Riley et al. (1995)	84 pacientes con diabetes mellitus tipo 1	Australia: Centro Menzies para la Investigación de Salud de la Población, Universidad de Tasmania. Los pacientes fueron seleccionados al azar de un registro de diabetes tratada con insulina basada en la población.	FFQ es un cuestionario utilizado para obtener información de frecuencia y en algunos casos, del tamaño de las porciones sobre el consumo de alimentos y bebidas durante un período específico de tiempo, generalmente el último mes o año. El cuestionario generalmente incluye preguntas sobre 80 a 120 artículos de alimentos y bebidas, aunque existen versiones más cortas o más largas del cuestionario. Tarda entre 30 y 60 minutos en completarse. Puede ser administrado por un entrevistador en caso de poca alfabetización, así como cuando se realiza en niños. En este artículo, la versión final del FFQ consta de 153 ítems alimentarios.	1-Validez de criterio 2-Sensibilidad al cambio	1-Inadecuado 2-Inadecuado	- El coeficiente de correlación entre el FFQ y un registro dietético pesado de 2 días para los mismos nutrientes varió de $r = 0.38$ para la ingesta de proteínas a $r = 0.60$ para la ingesta de grasas saturadas. El coeficiente de correlación entre el FFQ y la ingesta dietética habitual real se estimó en $r = 0.60$ para la ingesta de grasas ajustadas en energía, $r = 0.36$ para la ingesta de proteínas ajustadas en energía, $r = 0.72$ para la ingesta de carbohidratos ajustada en energía, $r = 0.55$ para la ingesta de grasas saturadas ajustada en energía y $r = 0.77$ para la ingesta de colesterol ajustada por energía.

Estudio (autor y año)	Poplación / tipo de DM	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Food Frequency Questionnaire (FFQ)						
Coulibaly et al. (2008)	57 pacientes con diabetes mellitus tipo 2	Mali: Servicios de atención primaria de salud. El Centre National de lutte contre le Diabète (CNLD)	En este artículo, la versión final del FFQ consta de 53 alimentos.	1-Validez de contenido 2- Prueba de hipótesis 3-Sensibilidad al cambio	1-Inadecuado 2-Inadecuado 3-Dudoso	- La correlación significativa entre la ingesta de proteínas estimada usando el FFQ y los recuerdos de 48 h fue $r = 0.63$
Hong et al. (2010)	85 pacientes con diabetes mellitus tipo 2	Corea: los pacientes fueron reclutados del Programa Nacional de Diabetes de Corea (KNDP)	En este artículo, la versión final del FFQ consta de 85 alimentos.	1-Validez de criterio 2-Sensibilidad al cambio	1-Inadecuado 2-Inadecuado	- El coeficiente de correlación fue más alto para la energía ($r = 0.740$; $P < 0.00$) y más bajo para el hierro ($r = 0.269$; $P < 0.05$). Los valores de Kappa para energía, carbohidratos, proteínas, grasas y calcio fueron 0.54, 0.37, 0.36, 0.46 y 0.19, respectivamente.
Aguiar et al. (2013)	88 pacientes con diabetes mellitus tipo 2	Sur de Brasil Hospital de Clínicas de Porto Alegre. Pacientes ambulatorios	En este artículo, la versión final del FFQ consta de 98 alimentos.	1-Validez de criterio 2-Sensibilidad al cambio	1-Inadecuado 2-Dudoso	- Existieron coeficientes de correlación significativos entre FFQ y WDR para la mayoría de los nutrientes. Además, los gráficos de Bland-Altman entre el FFQ y el WDR para la ingesta de energía y macronutrientes mostraron que el FFQ puede usarse como un método alternativo al WDR.

Estudio (autor y año)	Poplación / tipo de DM	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Food Frequency Questionnaire (FFQ)						
Luevano-Contreras et al. (2013)	30 pacientes con diabetes mellitus tipo 2	Estados Unidos: Universidad de Illinois	En este artículo, la versión final del FFQ consta de 90 alimentos.	1-Fiabilidad 2- Validez de criterio 3-Sensibilidad al cambio	1-Inadecuado 2-Inadecuado 3-Inadecuado	- ICC = 0.98 - El coeficiente de correlación para este estudio entre FR y FFQ-T1 fue $r = 0.68$ y entre FR y FFQ-T2 fue $r = 0.8$.
Farukuoye et al. (2014)	27 familiares no diabéticos de pacientes con DM2, 66 pacientes con diabetes y 30 individuos sanos.	Viena: Servicio de Diabetes para pacientes ambulatorios. Hospital Hanusch, Hospital Docente y un Servicio de Fisioterapia local.	En este artículo, la versión final del FFQ consta de 107 alimentos.	1-Validez de criterio 2-Sensibilidad al cambio	1-Inadecuado 2-Inadecuado	- La correlación significativa entre FFQ y 7DR fue de 0.23-0.72
Liese et al. (2014)	172 pacientes con diabetes mellitus tipo 1	Estados Unidos: Centro de Investigación de Obesidad Nutricional de la Universidad de Carolina del Norte.	En este artículo, la versión final del FFQ consta de 85 alimentos.	1-Fiabilidad 2- Prueba de hipótesis 3- Sensibilidad al cambio	1-Inadecuado 2-Dudoso 3-Dudoso	- ICC ambos FFQ, $r = 0.24-0.71$. - Las correlaciones medias de los grupos de alimentos y nutrientes entre el FFQ y la ingesta habitual verdadera fueron $r = 0.41$ y $r = 0.38$, respectivamente.
Petersen et al. (2015)	67 pacientes con diabetes mellitus tipo 1 y 2	Australia: Universidad del sur de Australia y el Royal Adelaide Hospital. Los pacientes fueron reclutados de la comunidad.	En este artículo, la versión final del FFQ consta de 74 alimentos.	1-Validez de criterio 2- Sensibilidad al cambio	1-Inadecuado 2-Inadecuado	- Hubo una correlación de débil a moderada entre la ingesta de nutrientes medida con los dos métodos (registro de alimentos con control y FFQ) y una correlación de moderada a alta para la ingesta de alimentos.

Estudio (autor y año)	Poplación / tipo de DM	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Food Frequency Questionnaire (FFQ)						
Sami et al. (2017)	132 pacientes con diabetes mellitus tipo 2	Arabia Saudita	En este artículo, la versión final del FFQ consta de 99 alimentos.	1-Validez de contenido 2-Validez estructural 3-Consistencia interna 4-Prueba de hipótesis 5- Sensibilidad al cambio	1-Dudoso 2- Dudoso 3-Dudoso 4-Inadecuado 5-Inadecuado	- El EFA resultó una solución de 5 factores. El factor 1 comprendía los ítems 2, 8, 12 y 13. El factor 2 comprendía los ítems 1, 3, 6 y 1. El factor 3 estaba compuesto por los ítems 9, 14 y 16. El factor 4 estaba compuesto por los ítems 4 y 15. El factor 5 estaba compuesto por los ítems 5, 7 y 11. Las direcciones de las cargas fueron positivas y libres de complejidad factorial. Los 5 factores fueron etiquetados como "selección de alimentos", "impacto en la salud", "elecciones saludables", "restricción de alimentos" y "categorización de alimentos". - α Cronbach = 0.782-0.908. - La detección de la correlación antiimagen mostró que todas las correlaciones diagonales estaban por encima del valor umbral de 0.50, el rango de correlaciones era de $r = 0.63 - 0.92$.

Estudio (autor y año)	Poplación / tipo de DM	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Food Frequency Questionnaire (FFQ)						
Meng-Chuan et al. (2018)	126 pacientes con diabetes mellitus tipo 2	Taiwán: Hospital de la Universidad de Medicina de Kaohsiung.	En este artículo, la versión final del FFQ consta de 45 alimentos.	1-Validez de criterio 2- Sensibilidad al cambio	1-Inadecuado 2-Inadecuado	- Los coeficientes de correlación de Pearson parciales derivados de los FFQ y tres 24-HDR y ajustados por energía fueron de $r = 0.65$, $r = 0.58$, $r = 0.64$ y $r = 0.66$ para proteínas, grasas, carbohidratos y fibra, respectivamente.
Perceived Dietary Adherence Questionnaire (PDAQ)						
Asaad et al. (2015)	73 pacientes con diabetes mellitus tipo 2	Canadá: los pacientes fueron reclutados de la comunidad	El Cuestionario de adherencia dietética percibida (PDAQ) consta de nueve preguntas con puntajes que van desde el 0, el más bajo al más alto 7 y se utiliza para medir las percepciones de los pacientes sobre su adherencia a la dieta. La respuesta se basa en una escala Likert de siete puntos. Los puntajes más altos reflejan una mayor adherencia, excepto los ítems 4 y 9, que indican elecciones poco saludables.	1-Validez de contenido 2-Consistencia interna 3- Fiabilidad 4- Prueba de hipótesis 5- Sensibilidad al cambio	1-Dudoso 2-Inadecuado 3-Inadecuado 4-Dudoso 5-Dudoso	- Coeficiente α de Cronbach = 0.78 - ICC = 0.78 - Los coeficientes de correlación para los ítems de PDAQ versus retiros de 24 h variaron de $r = 0.46$ a $r = 0.11$

Estudio (autor y año)	Poplación / tipo de DM	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
UK Diabetes and Diet Questionnaire (UKDDQ)						
England et al. (2016)	177 pacientes con diabetes mellitus tipo 2	Reino Unido (suroeste de Inglaterra): los pacientes fueron reclutados en el programa de salud STAMP 2: tiempo sedentario y salud metabólica en personas con (o en riesgo de) diabetes tipo 2	El UK Diabetes and Diet Questionnaire (UKDDQ) se compone de 25 elementos, de esos 25 elementos, 20 contribuyen a la puntuación general. Cada uno de los 20 ítems tiene seis categorías para que el participante elija lo que corresponde a la frecuencia de consumo del participante para ese ítem en particular, el puntaje para cada ítem varía de 0 a 5 (siendo 0 más saludable y 5 menos saludable).	1-Validez de contenido 2- Fiabilidad 3- Prueba de hipótesis 4- Sensibilidad al cambio	1-Inadecuado 2-Dudoso 3-Dudoso 4-Inadecuado	- ICC = 0.90 - Los puntajes totales del UKDDQ y los diarios de alimentos mostraron similitudes ICC = 0.54. La correlación para ítems individuales varió y fue buena para "alcohol" ($\kappa_w = 0.71$) y "cereales para el desayuno" ($\kappa_w = 0.70$), no hubo correlación para "vegetales" ($\kappa_w = 0.08$) o "pasteles salados" ($\kappa_w = 0.09$).

Estudio (autor y año)	Poplación / tipo de DM	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
The self-developed Dietary Knowledge Questionnaire (DKQ/DK)						
Sami et al. (2017)	132 pacientes con diabetes mellitus tipo 2	Arabia Saudita	El cuestionario de conocimiento dietético (DKQ/DK) consta de 20 preguntas de opción múltiple. Las preguntas evalúan el conocimiento dietético de un paciente para carbohidratos, lípidos, proteínas, tipo de alimento y opciones de alimentos. Cada pregunta de opción múltiple posee solo una respuesta correcta. Las respuestas están codificadas como 1 = correcto y 0 = incorrecto, y no lo sé. El puntaje varía de 0 a 20; un nivel de conocimiento dietético más alto se indica mediante una puntuación más alta. Se considera que un puntaje <50% presenta un conocimiento dietético pobre, un puntaje entre 50% -75% obtiene un buen conocimiento dietético y un puntaje > 75% se considera que tiene un conocimiento dietético adecuado.	1-Validez de contenido 2-Validez estructural 3-Consistencia interna 4-Prueba de hipótesis 5-Sensibilidad al cambio	1-Dudoso 2-Dudoso 3-Dudoso 4-Inadecuado 5-Inadecuado	- El EFA resultó una solución de 5 factores. El factor 1 comprendía los ítems 2, 8, 12 y 13. El factor 2 comprendía los ítems 1, 3, 6 y 1. El factor 3 estaba compuesto por los ítems 9, 14 y 16. El factor 4 estaba compuesto por los ítems 4 y 15. El factor 5 estaba compuesto por los ítems 5, 7 y 11. Las direcciones de las cargas fueron positivas y libres de complejidad factorial. Los 5 factores fueron etiquetados como "selección de alimentos", "impacto en la salud", "elecciones saludables", "restricción de alimentos" y "categorización de alimentos". - α Cronbach = 0.869. - La correlación total para DKQ varió de $r = 0.364 - 0.588$. Todos los valores estaban por encima del valor de corte > 0.30.

Estudio (autor y año)	Poplación / tipo de DM	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Diabetes Mellitus Knowledge Questionnaire (DMK)						
Sami et al. (2017)	132 pacientes con diabetes mellitus tipo 2	Arabia Saudita	El cuestionario DMK utilizado en esta investigación es la versión modificada / extendida del cuestionario DKQ desarrollado por Garcia et al. La versión original consta de 24 preguntas. La nueva versión del cuestionario DMK utilizada en esta investigación consta de 30 preguntas que se utilizan para evaluar con mayor detalle la "comprensión de la diabetes, los síntomas y el conocimiento, las complicaciones, la alimentación y el ejercicio, la medicación, la prevención y las investigaciones". Las respuestas se codifican como 1 = sí y 0 = no, y no lo sé.	1-Validez de contenido 2-Validez estructural 3-Consistencia interna 4-Prueba de hipótesis 5- Sensibilidad al cambio	1-Dudoso 2-Dudoso 3-Dudoso 4-Inadecuado 5-Inadecuado	- El EFA resultó una solución de 5 factores. El factor 1 comprendía los ítems 2, 8, 12 y 13. El factor 2 comprendía los ítems 1, 3, 6 y 1. El factor 3 estaba compuesto por los ítems 9, 14 y 16. El factor 4 estaba compuesto por los ítems 4 y 15. El factor 5 estaba compuesto por los ítems 5, 7 y 11. Las direcciones de las cargas fueron positivas y libres de complejidad factorial. Los 5 factores fueron etiquetados como "selección de alimentos", "impacto en la salud", "elecciones saludables", "restricción de alimentos" y "categorización de alimentos". - α Cronbach = 0.891. - La correlación total para DMK varió de $r = 0.358 - 0.529$. Todos los valores estaban por encima del valor de corte > 0.30 .

Estudio (autor y año)	Poplación / tipo de DM	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Motiv.Diaf-DM2 Questionnaire (MDDM2)						
Martín Payo et al. (2018)	206 pacientes con diabetes mellitus tipo 2	España: servicios de atención primaria	Motiv. Diaf-DM2 se compone de tres bloques. Incluye variables sociodemográficas, preguntas relacionadas con el tipo de motivación de los pacientes en relación con la realización de AF y resiliencia. Este cuestionario consta de 4 ítems en formato Likert en una escala de 1 (no me describe en absoluto) a 5 (me describe muy bien), siendo el puntaje más alto el comportamiento más resistente.	1-Validez de contenido 2-Validez estructural 3-Consistencia interna 4- Fiabilidad 5- Prueba de hipótesis 6-Sensibilidad al cambio	1-Inadecuado 2-Inadecuado 3-Inadecuado 4-Inadecuado 5-Dudoso 6-Dudoso	- La estructura de la prueba es bidimensional. Primero un análisis descriptivo y posteriormente un EFA. - α Cronbach = 0,756 para el primer factor y α Cronbach = 0,821 para el segundo factor. - Correlación de Pearson = 0.604 y 0.638 para cada factor, respectivamente. -La resiliencia tuvo una correlación estadísticamente significativa con NB dietético ($r = 0.398$; $p < 0.001$) y NB de actividad física ($r = 0.505$; $p < 0.001$).

Abreviaturas: (DM) Diabetes Mellitus; (DKB) Diabetes Knowledge and Behaviour Questionnaire; (FFQ) Food Frequency Questionnaire; (CNLD) Centre National de lutte contre le Diabète ; (KNDP) Korean National Diabetes Program; (WDR) Weighed Diet Record; (ICC) Intraclass Correlation Coefficient; (7DR) 7-day food record; (EFA) Análisis Factorial Exploratorio; (24-HDR) 24-hours dietary recalls; (PDAQ) Perceived Dietary Adherence Questionnaire; (UKDDQ) United Kingdom Diabetes and Diet Questionnaire; (DKQ) The self-developed Dietary Knowledge Questionnaire; (DK) Dietary Knowledge; (DMK) Diabetes Mellitus Knowledge Questionnaire; (MDDM2) Motiv.Diaf-DM2 Questionnaire; (AF) Actividad Física; (NB) Necesidades Básicas Psicológicas; (HbA1c) Hemoglobina glicosilada.

3.1 Propiedades psicométricas

Las principales propiedades evaluadas fueron validez de contenido, validez de criterio, prueba de hipótesis y sensibilidad al cambio. Los resultados de la lista de verificación COSMIN de cuatro puntos de los instrumentos para medir el estado nutricional en pacientes con DM1 y/o DM2 se muestran en la tabla 3.

Tabla 3. Calidad metodológica de los estudios y calidad de los resultados informados por propiedad de medición, instrumento y estudio.

Instrumento	Artículo	Validez de contenido		Validez estructural		Consistencia interna		Fiabilidad		Validez de criterio		Prueba de hipótesis		Sensibilidad al cambio	
		M	Q	M	Q	M	Q	M	Q	M	Q	M	Q	M	Q
DKB	Simmons et al. (1994)					I	-					D	?	D	?
	Riley et al. (1995)									I	-			I	+
	Coulibaly et al. (2008)	I	?									I	?	D	?
	Hong et al. (2010)									I	-			I	+
	Aguiar et al. (2013)									I	-			D	+
FFQ	Luevano-Contreras et al. (2013)							I	+	I	-			I	+
	Farukuoye et al. (2014)									I	-			I	+
	Liese et al. (2014)							I	-			D	?	D	?
	Petersen et al. (2015)									I	+			I	+
	Sami et al. (2017)	D	-	D	?	D	+					I	+	I	+
	Meng-Chuan et al. (2018)									I	-			I	+
PDAQ	Asaad et al. (2015)	D	-			I	+	I	+			D	?	D	?
UKDDQ	England et al. (2016)	I	+					D	+			D	?	I	?
			-												
DKQ/DK	Sami et al. (2017)	D	-	D	?	D	+					I	+	I	+
DMK	Sami et al. (2017)	D	-	D	?	D	+					I	+	I	+
MDDM2	Martín Payo et al. (2018)	I	?	I	-	I	+	I	-			D	?	D	?

La validez transcultural / invariancia de medición y el error de medición se omitieron en la tabla porque no se evaluaron en ningún estudio.

Q: Calidad metodológica del estudio calificado como I = Inadecuado, D = Dudoso, A = Adecuado, V = Muy bueno, casillas vacías = no informado.

P: Calidad de los resultados calificados como + = calificación suficiente, ? = calificación indeterminada, ± = calificación inconsistente, - = calificación insuficiente, casillas vacías = no informado

La tabla 4 muestra datos sobre los instrumentos para medir la ingesta alimentaria utilizando la lista de verificación COSMIN y el método GRADE.

Tabla 4. Resumen de hallazgos por propiedad de medición e instrumento.

Instrumento	Validez de contenido			Validez estructural			Consistencia interna			Fiabilidad			Validez de criterio			Prueba de hipótesis			Sensibilidad al cambio		
	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE
DKB							I	-	VL							D	?	L	D	?	L
FFQ	I	-	VL	D	?	VL	D	+	L	I	-	VL	I	-	L	I	?	L	I	?	L
PDAQ	D	-	L				I	+	VL	I	+	VL				D	?	L	D	?	L
UKDDQ	I	+	VL							D	+	L				D	?	L	I	?	VL
		-																			
DKQ/DK	D	-	L	D	?	L	D	+	L							I	+	VL	I	+	VL
DMK	D	-	L	D	?	L	D	+	L							I	+	VL	I	+	VL
MDDM2	I	?	VL	I	-	VL	I	+	VL	I	-	VL				D	?	L	D	?	L

La validez transcultural / invariancia de medición y error de medición se omitieron en la tabla porque no se evaluaron en ningún estudio.

M: Calidad metodológica del estudio calificado como I = Inadecuado, D = Dudoso, A = Adecuado, V = Muy bueno, casillas vacías = no informado

Q: Calidad de los resultados calificados como + = calificación suficiente, ? = calificación indeterminada, ± = calificación inconsistente, - = calificación insuficiente, casillas vacías = no informado.

QE: Calidad de la evidencia calificada como H = Alta, M = Moderada, L = Baja, VL = Muy baja.

3.1.1 *Diabetes Knowledge and Behaviour Questionnaire (DKB)*

Las propiedades psicométricas de este cuestionario fueron evaluadas en un solo estudio (30). La puntuación para consistencia interna fue “inadecuada”, con un alfa de Cronbach entre 0.59 y 0.90. La prueba de hipótesis se calificó como “dudosa”, puesto que el cuestionario de frecuencia alimentaria de 12 ítems (DKB) mostró valores de correlación moderados con la evaluación dietética de las calorías totales ($r = 0.48 - 0.64$) y de las calorías debidas a la grasa ($r = 0.41 - 0.65$), y existió una correlación negativa entre una pregunta simple relacionada con la frecuencia del consumo de fruta y el índice de grasa en europeos ($r = 0.25$) y maoríes ($r = -0.33$).

3.1.2 *Food Frequency Questionnaire (FFQ)*

Este cuestionario fue evaluado en 10 estudios (31-40). Únicamente en 2 de esos estudios (33,40) se valoró la validez de contenido, puntuando entre “inadecuada” y “dudosa”, debido principalmente a la falta de valoración de esta propiedad por parte de profesionales y de pacientes con diabetes. La validez estructural fue evaluada en el estudio de Sami et al. (2017) (40), obteniendo una puntuación “dudosa”, y se realizó un análisis factorial exploratorio (EFA) que dio como resultado una solución de 5 factores.

El cuestionario obtuvo un alfa de Cronbach entre 0.782 y 0.908 (40). La fiabilidad test-retest se midió para 2 estudios, resultando dicha puntuación como “inadecuada” y siendo sus valores ICC=0.98 (36) e ICC entre 0.24 y 0.71 (37), respectivamente. En la mayoría de los estudios fue comparado el FFQ con un “gold standard”, un registro dietético pesado (Weighed Dietary Record), y la puntuación reflejada en validez de criterio resultó como “inadecuada” para la totalidad de dichos estudios (31,32,34,35,37-39), las correlaciones mostradas variaron entre $r=0.23$ (34) y $r=0.8$ (37).

La prueba de hipótesis resultó entre “inadecuada” y “dudosa”. Para el estudio de Coulibaly et al. (2008) (33), existió una correlación significativa entre la ingesta de proteínas estimada usando el FFQ y los retiros de 48 h ($r = 0.63$); para Liese et al. (2014) (36), las correlaciones medias, ajustadas por error de medición de los grupos de alimentos y nutrientes entre el FFQ y la ingesta habitual

verdadera fueron $r = 0.41$ y $r = 0.38$, respectivamente; finalmente para Sami et al. (2017) (40), las correlaciones entre los diferentes ítems del FFQ oscilaron entre 0.632 - 0.928.

3.1.3 *Perceived Dietary Adherence Questionnaire (PDAQ)*

El PDAQ solo fue descrito en un estudio (41), en el cual se evaluó la validez de contenido, calificándose como "dudosa". El alfa de Cronbach fue 0.78 y el ICC=0.78. Para la prueba de hipótesis, los coeficientes de correlación para ítems PDAQ versus retiros de 24 h variaron de $r = 0.46$ a $r = 0.11$.

3.1.4 *UK Diabetes and Diet Questionnaire (UKDDQ)*

UKDDQ fue evaluado en un estudio (42), en el cual la validez de contenido puntuó como "inadecuada" debido a la valoración de dicha propiedad por parte de profesionales y pacientes con diabetes. En cuanto a su fiabilidad, se obtuvo un ICC=0.90. Para la prueba de hipótesis existieron correlaciones moderadas entre los puntajes totales del UKDDQ y los diarios de alimentos en cuanto al consumo de alcohol ($\kappa_w = 0.71$) y "cereales para el desayuno" ($\kappa_w = 0.70$).

3.1.5 *The self-developed Dietary Knowledge Questionnaire (DKQ)*

Un solo estudio valoró las propiedades psicométricas de este cuestionario (40). La validez de contenido fue calificada como "dudosa", al igual que la validez estructural, donde se realizó un EFA, dando como resultado una solución de 5 factores. Estos 5 factores fueron etiquetados de la siguiente forma: "selección de alimentos", "impacto en la salud", "elecciones saludables", "restricción de alimentos" y "categorización de alimentos". En dicho estudio, la consistencia interna obtuvo un valor de alfa de Cronbach de 0.869. Para la prueba de hipótesis, la correlación total para DKQ entre sus diferentes ítems varió de $r = 0.364$ - 0.588.

3.1.6 *Diabetes Mellitus Knowledge Questionnaire (DMK)*

DMK fue evaluado para un único estudio (40), resultando como "dudosa" la validez de contenido, al igual que la validez estructural, donde se llevó a cabo

un EFA, obteniéndose una solución de 5 factores. Estos 5 factores fueron etiquetados de la siguiente forma: "selección de alimentos", "impacto en la salud", "elecciones saludables", "restricción de alimentos" y "categorización de alimentos". Para este estudio, el alfa de Cronbach fue 0.891. En la prueba de hipótesis, la correlación total para DMK entre sus diferentes ítems varió de $r = 0.358 - 0.529$

3.1.7 *Motiv.Diaf-DM2 Questionnaire (MDDM2)*

El cuestionario fue incluido en uno solo de los estudios (43). En este, tanto la validez de contenido como la validez estructural fueron calificadas como "inadecuadas". Este cuestionario mostró una estructura bidimensional (ingesta dietética y adherencia a la actividad física). El estudio obtuvo un alfa de Cronbach de 0.756 para el primer factor y 0.821 para el segundo factor. Para la prueba de hipótesis, la primera dimensión no obtuvo correlaciones significativas con las variables evaluadas (necesidades psicológicas básicas, resiliencia y hemoglobina glicosilada (HbA1c)). Las puntuaciones para la segunda dimensión (adherencia a la actividad física) estaban en consonancia con las necesidades psicológicas básicas ($r = 0.281$) y la resistencia ($r = 0.216$). Además, se encontró una relación relativamente moderada con la HbA1c ($r = -0.182$).

4. DISCUSIÓN

Se trata de la primera revisión sistemática en la que se identifican instrumentos para evaluar el estado nutricional en pacientes con DM1 o DM2, y se evalúan sus propiedades psicométricas y la calidad metodológica basándose en COSMIN (19,20). En este estudio se identificaron 7 instrumentos que fueron diseñados específicamente para evaluar el estado nutricional en pacientes diabéticos a través de la inclusión de 14 estudios.

Diferentes factores como el ámbito, la población donde se utilizará el instrumento, sus dimensiones, el número de ítems y la evidencia mostrada en la evaluación de cada propiedad psicométrica deben ser considerados por clínicos e investigadores para decidir qué instrumento es el mejor para medir el estado nutricional. Además, se debe tener en cuenta el idioma y la cultura de la versión original del instrumento. Se han publicado revisiones sistemáticas sobre el estado

nutricional enfocadas a ancianos y niños, pero estos instrumentos no han sido validados específicamente para personas con DM1 o DM2; el uso de estos instrumentos en esta población podría darnos información diferente sobre su estado nutricional (44,45).

El cuestionario FFQ incluido en ambas revisiones sistemáticas (44,45) ha sido validado en personas con DM1 o DM2, y sus propiedades psicométricas evaluadas en el presente estudio (31–40). No se contó con un instrumento de medición del estado nutricional para el cual se evaluaran todas las propiedades psicométricas. La validez transcultural y el error de medición fueron las únicas 2 propiedades que no estuvieron presentes en ninguno de los 7 instrumentos seleccionados para la revisión, lo que refleja las limitaciones en este tipo de herramientas. Las propiedades psicométricas evaluadas en la mayoría de los estudios fueron la validez de contenido y de criterio, la prueba de hipótesis y la sensibilidad al cambio. La validez de criterio se evaluó completamente para el FFQ porque en 7 de los 10 estudios se comparó con un "gold standard" (31,32,34,35,37–39).

La calidad metodológica de los estudios para cada propiedad psicométrica evaluada no fue la esperada, siendo "inadecuada" o "dudosa" para todos los instrumentos, debido en el caso de la validez de contenido a la falta de evaluación de la relevancia de los ítems en la población de estudio. En cuanto a la consistencia interna, no se realizó análisis factorial ni se calculó la consistencia interna para cada dimensión por separado, faltó la descripción de los porcentajes y el manejo de los valores faltantes. Para la prueba de hipótesis, las consecuencias de las hipótesis no estaban claras o no estaban formuladas *a priori* (aunque era posible deducir lo que se esperaba) y faltaba la descripción de la gestión de los valores faltantes.

Finalmente, para la validez estructural, la fiabilidad y la validez de criterio, faltó la descripción de los porcentajes y el manejo de los valores perdidos. Las diferencias en los valores de las propiedades psicométricas de los instrumentos incluidos podrían explicarse por el hecho de que, en cada estudio, el ámbito y las características de la población fueron distintas. A pesar de estas desigualdades, los diferentes cuestionarios incluían diversas preguntas sobre la ingesta dietética que podrían influir en los valores de fiabilidad y validez.

DKB (30), PDAQ (41), UKDDQ (42), DKQ (40), DMK (40) y MDDM2 (43) solo se evaluaron en un estudio. Se requieren estudios adicionales para evaluar las propiedades psicométricas de estos instrumentos.

La calidad de la evidencia fue "baja" o "muy baja" para todas las propiedades de medición de los instrumentos incluidos en esta revisión sistemática. La validez de contenido, la validez estructural y la consistencia interna se consideran las propiedades de medición más importantes (19,21). La validez de contenido estuvo presente en 6 de los 7 instrumentos (FFQ (33,40), PDAQ (41), UKDDQ (42), DKQ, DMK (40) y MDDM2 (43)), y para todos ellos la calidad de la evidencia fue "baja" o "muy baja"; excepto para UKDDQ (42), en la que se calificó como "±", lo que indica que los elementos incluidos son inconsistentes o indeterminados con respecto al constructo que se pretendía medir. Con respecto a la validez estructural, nuevamente la calidad de la evidencia fue "baja" o "muy baja" para 4 de los instrumentos: FFQ, DKQ, DMK (40) y MDDM2 (43). Se obtuvo una estructura de 5 factores para FFQ, DKQ y DMK (40), y en el caso de MDDM2 (43) solo hubo 2 factores resultantes.

Finalmente, se obtuvo una calificación de "baja" o "muy baja" para todos los instrumentos que evaluaron la consistencia interna. En la mayoría de las herramientas utilizadas para evaluar el estado nutricional en pacientes con diabetes (FFQ (40), PDAQ (41), DKQ, DMK (40) y MDDM2 (43)), la calidad de los resultados se calificó como "+" (alfa de Cronbach ≥ 0.70), mientras que para DKB (30) fue "-" (alfa de Cronbach < 0.70).

FFQ (31–40) ha sido el cuestionario incluido en mayor volumen de estudios (10 estudios en total), en el que más características psicométricas se han evaluado y con una mejor puntuación en cuanto a la calidad de la evidencia. Este cuestionario ha sido adaptado a personas con DM1 o DM2, incluyendo alimentos específicos que se deben medir en esta población. Esto hace que este cuestionario sea el más ampliamente validado y recomendado para evaluar la ingesta dietética en este tipo de población.

4.1 Fortalezas y limitaciones

Esta es la revisión sistemática más actual para identificar, evaluar y resumir la evidencia sobre instrumentos que miden el estado nutricional en pacientes con

DM1 o DM2. La revisión se realizó de acuerdo con los estándares COSMIN, lo que asegura que se utilizó un método adecuado siguiendo las recomendaciones de los expertos (19,21). Aunque se efectuó una búsqueda minuciosa en varias bases de datos es posible que en esta revisión no se hayan identificado todos los instrumentos disponibles para la medición del estado nutricional debido a la limitación de la estrategia de búsqueda a solo artículos publicados en inglés y español, por lo que es probable que se haya prescindido de trabajos de investigación sobre instrumentos desarrollados y empleados en otros idiomas.

5. CONCLUSIONES

Se han desarrollado varios instrumentos para la evaluación de la ingesta dietética en personas con diabetes. La evaluación de este constructo es de gran utilidad, tanto en la práctica clínica como en la investigación, requiriendo nuevos conocimientos en esta área. La evaluación de la ingesta dietética es fundamental porque permite identificar y evaluar intervenciones enfocadas a mejorar la nutrición en personas con diabetes. FFQ es el mejor instrumento disponible para evaluar la ingesta dietética en personas con diabetes y cuenta con diferentes versiones en número de ítems. Es necesario un mayor número de estudios con una alta calidad metodológica, en los que se analicen todas las propiedades psicométricas de los instrumentos existentes y sean adaptados a otras culturas.

REFERENCIAS BIBLIOGRÁFICAS

1. Castro-Rodríguez, M.; Carnicero, J.A.; Garcia-Garcia, F.J.; Walter, S.; Morley, J.E.; Rodríguez-Artalejo, F.; Sinclair, A.J.; Rodríguez-Mañas, L. Frailty as a major factor in the increased risk of death and disability in older people with diabetes. *Journal of the American Medical Directors Association* 2016, 17, 949–955.
2. Saeedi, P.; Salpea, P.; Karuranga, S.; Petersohn, I.; Malanda, B.; Gregg, E.W.; Unwin, N.; Wild, S.H.; Williams, R. Mortality attributable to diabetes in 20–79 years old adults, 2019 estimates: Results from the International Diabetes Federation Diabetes Atlas, 9th edition. *Diabetes Research and Clinical Practice* 2020, 108086, doi: 10.1016/j.diabres.2020.108086.
3. Sun, Y.; You, W.; Almeida, F.; Estabrooks, P.; Davy, B. The effectiveness and cost of lifestyle interventions including nutrition education for diabetes prevention: a systematic review and meta-analysis. *Journal of the Academy of Nutrition and Dietetics* 2017, 117, 404–421.
4. Wylie-Rosett, J.; Herman, W.H.; Goldberg, R.B. Lifestyle intervention to prevent diabetes: intensive and cost effective. *Current opinion in lipidology* 2006, 17, 37–44.
5. Patel, Y.R.; Gadiraju, T.V.; Gaziano, J.M.; Djoussé, L. Adherence to healthy lifestyle factors and risk of death in men with diabetes mellitus: The Physicians' Health Study. *Clinical nutrition* 2018, 37, 139–143.
6. Williams, J.L.S.; Walker, R.J.; Smalls, B.L.; Campbell, J.A.; Egede, L.E. Effective interventions to improve medication adherence in Type 2 diabetes: a systematic review. *Diabetes management (London, England)* 2014, 4, 29.
7. Vasilevskis, E.E. The Challenges of Diabetes Medication Management at Hospital Discharge in Older Adults. *JAMA Network Open* 2020, 3, e201500–e201500.

8. Deng, A.; Pattanaik, S.; Bhattacharya, A.; Yin, J.; Ross, L.; Liu, C.; Zhang, J. Fish consumption is associated with a decreased risk of death among adults with diabetes: 18-year follow-up of a national cohort. *Nutrition, Metabolism and Cardiovascular Diseases* 2018, 28, 1012–1020.
9. Bonaccio, M.; Di Castelnuovo, A.; Costanzo, S.; Persichillo, M.; De Curtis, A.; Donati, M.B.; De Gaetano, G.; Iacoviello, L.; Investigators, M.-S. Adherence to the traditional Mediterranean diet and mortality in subjects with diabetes. Prospective results from the MOLI-SANI study. *European journal of preventive cardiology* 2016, 23, 400–407.
10. Jalilian, H.; Pezeshki, M.Z.; Janati, A.; Najafipour, F.; Sarbakhsh, P.; Zarnaq, R.K. Readiness for weight change and its association with diet knowledge and skills, diet decision making and diet and exercise barriers in patients with type 2 diabetes. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews* 2019, 13, 2889–2895.
11. Sami, W.; Alabdulwahhab, K.M.; Ab Hamid, M.R.; Alasbali, T.A.; Alwadani, F.A.; Ahmad, M.S. Dietary Knowledge among Adults with Type 2 Diabetes—Kingdom of Saudi Arabia. *International Journal of Environmental Research and Public Health* 2020, 17, 858, doi:10.3390/ijerph17030858.
12. Garcia-Molina, L.; Lewis-Mikhael, A.-M.; Riquelme-Gallego, B.; Cano-Ibáñez, N.; Oliveras-López, M.-J.; Bueno-Cavanillas, A. Improving type 2 diabetes mellitus glycaemic control through lifestyle modification implementing diet intervention: a systematic review and meta-analysis. *European journal of nutrition* 2019, 1–16.
13. Mozaffari, H.; Daneshzad, E.; Azadbakht, L. Dietary carbohydrate intake and risk of bone fracture: a systematic review and meta-analysis of observational studies. *Public Health* 2020, 181, 102–109.
14. Rodríguez-Monforte, M.; Flores-Mateo, G.; Barrio, F.; Costa, B.; Sánchez, E. Metabolic syndrome and dietary patterns: a systematic review and meta-analysis of observational studies—reply. *European journal of nutrition* 2019, 58, 3383–3386.

15. Knüppel, S.; Norman, K.; Boeing, H. Is a single 24-hour dietary recall per person sufficient to estimate the population distribution of usual dietary intake? *The Journal of nutrition* 2019, 149, 1491–1492.
16. Freedman, L.S.; Midthune, D.; Arab, L.; Prentice, R.L.; Subar, A.F.; Willett, W.; Neuhouser, M.L.; Tinker, L.F.; Kipnis, V. Combining a food frequency questionnaire with 24-hour recalls to increase the precision of estimation of usual dietary intakes—Evidence from the validation studies pooling project. *American journal of epidemiology* 2018, 187, 2227–2232.
17. Delichatsios, H.K.; Hunt, M.K.; Lobb, R.; Emmons, K.; Gillman, M.W. EatSmart: Efficacy of a multifaceted preventive nutrition intervention in clinical practice. *Preventive Medicine* 2001, 33, 91–98, doi:10.1016/S0091-7435(01)80004-9.
18. Patton, S.R.; Dolan, L.M.; Powers, S.W. Parent report of mealtime behaviors in young children with type 1 diabetes mellitus: Implications for better assessment of dietary adherence problems in the clinic. *Journal of Developmental & Behavioral Pediatrics* 2006, 27, 202–208.
19. Mokkink, L.B.; Terwee, C.B.; Patrick, D.L.; Alonso, J.; Stratford, P.W.; Knol, D.L.; Bouter, L.M.; de Vet, H.C.W. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010, 19, 539–549, doi:10.1007/s11136-010-9606-8.
20. Terwee, C.B.; Mokkink, L.B.; Knol, D.L.; Ostelo, R.W.J.G.; Bouter, L.M.; de Vet, H.C.W. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res* 2012, 21, 651–657, doi:10.1007/s11136-011-9960-1.
21. Prinsen, C.A.C.; Mokkink, L.B.; Bouter, L.M.; Alonso, J.; Patrick, D.L.; de Vet, H.C.W.; Terwee, C.B. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018, 27, 1147–1157, doi:10.1007/s11136-018-1798-3.

22. Guyatt, G.H.; Oxman, A.D.; Vist, G.E.; Kunz, R.; Falck-Ytter, Y.; Alonso-Coello, P.; Schünemann, H.J. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008, 336, 924–926, doi:10.1136/bmj.39489.470347.AD.
23. Han, C.Y.; Zheng, X.; Lee, L.F.; Chan, C.G.B.; Lee, Y.Q.; Zailani, N.A.; Ng, K.; Bhaskaran, K. Development of a diabetes-related nutrition knowledge questionnaire for individuals with type 2 diabetes mellitus in Singapore. *Nutrition & Dietetics* 2019, 76, 567–573, doi:10.1111/1747-0080.12513.
24. Quick, V.; Lipsky, L.M.; Nansel, T.R. Psychometric properties and factor structure of the adapted Self-Regulation Questionnaire assessing autonomous and controlled motivation for healthful eating among youth with type 1 diabetes and their parents. *Child Care Health Dev* 2018, 44, 651–658, doi:10.1111/cch.12574.
25. Samuel-Hodge, C.D.; DeVellis, R.F.; Ammerman, A.; Keyserling, T.C.; Elasy, T.A. Reliability and validity of a measure of perceived diabetes and dietary competence in African American women with type 2 diabetes. *The Diabetes Educator* 2002, 28, 979–988.
26. Rohani, H.; Eslami, A.A.; Ghaderi, A. Development and psychometric evaluation of a health action process approach inventory for healthful diet among type 2 diabetes patients. *International journal of preventive medicine* 2016, 7.
27. Anastasiou, C.A.; Fappa, E.; Zachari, K.; Mavrogianni, C.; Van Stappen, V.; Kivelä, J.; Virtanen, E.; González-Gil, E.M.; Flores-Barrantes, P.; Nánási, A. Development and reliability of questionnaires for the assessment of diet and physical activity behaviors in a multi-country sample in Europe the Feel4Diabetes Study. *BMC endocrine disorders* 2020, 20, 1–7.
28. Brito-Brito, P.R.; Martín-García, Á.; Oter-Quintana, C.; Paloma-Castro, O.; Romero-Sánchez, J.M. Development and Content Validation of a NOC-Based Instrument for Measuring Dietary Knowledge in Patients with

- Diabetes: CoNOCidiet-Diabetes. *International journal of nursing knowledge* 2019.
29. Stetson, B.; Schlundt, D.; Rothschild, C.; Floyd, J.E.; Rogers, W.; Mokshagundam, S.P. Development and validation of The Personal Diabetes Questionnaire (PDQ): a measure of diabetes self-care behaviors, perceptions and barriers. *Diabetes research and clinical practice* 2011, 91, 321–332.
 30. Simmons, D.; Mandell, C.; Fleming, C.; Gatland, B.; Leakehe, L. Evaluation of a diabetes knowledge and behaviour (DKB) questionnaire. *Asia Pac J Clin Nutr* 1994, 3, 193–200.
 31. Huang, M.-C.; Lin, K.-D.; Chen, H.-J.; Wu, Y.-J.; Chang, C.-I.; Shin, S.-J.; Hung, H.-C.; Lee, C.-H.; Huang, Y.-F.; Hsu, C.-C. Validity of a Short Food Frequency Questionnaire Assessing Macronutrient and Fiber Intakes in Patients of Han Chinese Descent with Type 2 Diabetes. *International Journal of Environmental Research and Public Health* 2018, 15, 1142, doi:10.3390/ijerph15061142.
 32. Sarmiento, R.A.; Antonio, J.P.; Riboldi, B.P.; Montenegro, K.R.; Friedman, R.; Azevedo, M.J. de; Almeida, J.C. de Reproducibility and validity of a quantitative FFQ designed for patients with type 2 diabetes mellitus from southern Brazil. *Public Health Nutrition* 2014, 17, 2237–2245, doi:10.1017/S1368980013002644.
 33. Coulibaly, A.; O'Brien, H.T.; Galibois, I. Validation of an FFQ to assess dietary protein intake in type 2 diabetic subjects attending primary health-care services in Mali. *Public Health Nutrition* 2009, 12, 644–650, doi:10.1017/S1368980008002620.
 34. Farukuoye, M.; Strassburger, K.; Kacerovsky-Bielesz, G.; Giani, G.; Roden, M. Validity and reproducibility of an interviewer-administered food frequency questionnaire in Austrian adults at risk of or with overt diabetes mellitus. *Nutrition Research* 2014, 34, 410–419, doi:10.1016/j.nutres.2014.04.004.

35. Hong, S.; Choi, Y.; Lee, H.-J.; Kim, S.H.; Oe, Y.; Lee, S.Y.; Nam, M.; Kim, Y.S. Development and Validation of a Semi-Quantitative Food Frequency Questionnaire to Assess Diets of Korean Type 2 Diabetic Patients. *Korean Diabetes Journal* 2010, 34, 32–39, doi:10.4093/kdj.2010.34.1.32.
36. Liese, A.D.; Crandell, J.L.; Toozee, J.A.; Fangman, M.T.; Couch, S.C.; Merchant, A.T.; Bell, R.A.; Mayer-Davis, E.J. Relative validity and reliability of an FFQ in youth with type 1 diabetes. *Public Health Nutrition* 2015, 18, 428–437, doi:10.1017/S1368980014000408.
37. Luevano-Contreras, C.; Durkin, T.; Pauls, M.; Chapman-Novakofski, K. Development, relative validity, and reliability of a food frequency questionnaire for a case-control study on dietary advanced glycation end products and diabetes complications. *International Journal of Food Sciences and Nutrition* 2013, 64, 1030–1035, doi:10.3109/09637486.2013.816939.
38. Petersen, K.S.; Smith, J.M.; Clifton, P.M.; Keogh, J.B. Dietary intake in adults with type 1 and type 2 diabetes: validation of the Dietary Questionnaire for Epidemiological Studies version 2 FFQ against a 3-d weighed food record and 24-h urinalysis. *British Journal of Nutrition* 2015, 114, 2056–2063, doi:10.1017/S0007114515003748.
39. Riley, M.D.; Blizzard, L. Comparative Validity of a Food Frequency Questionnaire for Adults with IDDM. *Diabetes Care* 1995, 18, 1249–1254, doi:10.2337/diacare.18.9.1249.
40. Sami, W.; Ansari, T.; Butt, N.S.; Ab Hamid, M.R. Psychometric evaluation of dietary habits questionnaire for type 2 diabetes mellitus. *J. Phys.: Conf. Ser.* 2017, 890, 012151, doi:10.1088/1742-6596/890/1/012151.
41. Asaad, G.; Sadegian, M.; Lau, R.; Xu, Y.; Soria-Contreras, D.C.; Bell, R.C.; Chan, C.B. The Reliability and Validity of the Perceived Dietary Adherence Questionnaire for People with Type 2 Diabetes. *Nutrients* 2015, 7, 5484–5496, doi:10.3390/nu7075231.

42. England, C.Y.; Thompson, J.L.; Jago, R.; Cooper, A.R.; Andrews, R.C. Development of a brief, reliable and valid diet assessment tool for impaired glucose tolerance and diabetes: the UK Diabetes and Diet Questionnaire. *Public Health Nutrition* 2017, 20, 191–199, doi:10.1017/S1368980016002275.
43. Martín Payo, R.; González Méndez, X.; Papín Cano, C.; Suarez Alvarez, J. Development and validation of a questionnaire for assessing the characteristics of diet and physical activity in patients with type 2 diabetes. *Psicothema*, 30 2018.
44. Donini, L.M.; Savina, C.; Rosano, A.; Cannella, C. Systematic review of nutritional status evaluation and screening tools in the elderly. *Journal of Nutrition Health and Aging* 2007, 11, 421.
45. Bell, L.K.; Golley, R.K.; Magarey, A.M. Short tools to assess young children's dietary intake: a systematic review focusing on application to dietary index research. *Journal of obesity* 2013, 2013.

Artículo 2

The dietary knowledge measurement instruments in diabetes: a systematic psychometric review

PUBLICADO EN JOURNAL OF ADVANCED NURSING: 9 DE FEBRERO DE 2021.

Journal of Advanced Nursing, 2021, vol. 77, no 6, p. 2595-2622; <https://doi.org/10.1111/jan.14762>

THE DIETARY KNOWLEDGE MEASUREMENT INSTRUMENTS IN DIABETES: A SYSTEMATIC PSYCHOMETRIC REVIEW

RESUMEN

Objetivos: Identificar, evaluar críticamente y resumir instrumentos para medir el conocimiento dietético en pacientes con diabetes mellitus tipo 1 o 2.

Diseño: Revisión sistemática psicométrica.

Fuentes de datos: Se realizó una búsqueda exhaustiva durante marzo de 2020 en las siguientes bases de datos: Medline, Web of Science, Open Grey, Cochrane Library, base de datos COSMIN y Scielo.

Metodología: Se llevó a cabo una revisión sistemática de las propiedades de medición de los instrumentos que evalúan el nivel de conocimiento dietético en la población diabética de acuerdo con los Estándares más actualizados basados en el Consenso para la selección de la metodología de INstrumentos de Medición del estado de salud (COSMIN) y Elementos de Informe Preferidos para Declaraciones Sistemáticas y Metanálisis (PRISMA). La lista de verificación COSMIN se utilizó para evaluar las propiedades de medición de las medidas de resultado existentes informadas por el paciente para seleccionar el instrumento de medición de resultado más apropiado. El protocolo de esta revisión se inscribió en el Registro Prospectivo Internacional de Revisiones Sistemáticas (PROSPERO) (CRD42020186516).

Resultados: Se identificaron 9 instrumentos. Michigan Diabetes Knowledge Test (DKT) fue el cuestionario incluido en el mayor volumen de estudios. Diabetes Knowledge Questionnaire (DKQ) fue el cuestionario con más características psicométricas evaluadas. Personal Diabetes Questionnaire (PDQ) fue el segundo con mayor número de características psicométricas evaluadas y con la mejor puntuación en cuanto a la calidad de la evidencia.

Conclusión: Según la calidad metodológica de los resultados; DKT, DKQ y PDQ son los instrumentos recomendados para evaluar el conocimiento dietético en personas con diabetes tipo 1 y 2.

Palabras clave: diabetes mellitus; conocimiento dietético; instrumentos; enfermería; medidas de resultado.

ABSTRACT

Aims: To identify, critically appraise, and summarize instruments to measure dietary knowledge in patients with diabetes mellitus type 1 or 2.

Design: Psychometric systematic review.

Data sources: A comprehensive search was carried out during March 2020 in the followed databases: Medline, Web of Science, Open Grey, Cochrane Library, COSMIN database and Scielo.

Review Methods: In accordance with the most up to date COnsensus-based Standards for the selection of health status Measurement INstruments methodology (COSMIN) and Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statements. A systematic review of the measurement properties of the instruments that assess the level of dietary knowledge in the diabetic population was carried out. The COSMIN checklist was used to assess the measurement properties of existing patient-reported outcome measures to select the most appropriate outcome measurement instrument. The protocol of this systematic review was registered in the International Prospective Register of Systematic Reviews (PROSPERO) (CRD42020186516).

Results: 9 instruments were identified. The Michigan Diabetes Knowledge Test (DKT) was the questionnaire included in the largest volume of studies. The Diabetes Knowledge Questionnaire (DKQ) was the questionnaire with the most evaluated psychometric characteristics. The Personal Diabetes Questionnaire (PDQ) was the second with the highest number of psychometric characteristics evaluated and with a better score regarding the quality of the evidence.

Conclusion: According to the quality methodological results; DKT, DKQ and PDQ are the recommended instruments to assess dietary knowledge in people with diabetes types 1 and 2.

Keywords: diabetes mellitus; dietary knowledge; instruments; nursing; outcome measures.

1. INTRODUCCIÓN

La diabetes mellitus (DM) es definida por la Organización Mundial de la Salud (OMS) como “un trastorno metabólico de etiología múltiple caracterizado por hiperglucemia crónica con alteraciones en el metabolismo de carbohidratos, grasas y proteínas como resultado de defectos en la secreción de insulina, acción de la insulina o ambos”(1).

En 2019, se descubrió que un total de 463 millones de personas vivían con DM, lo que representa el 9.3% de la población adulta mundial (20 a 79 años). Se espera que este número aumente a 578 millones en 2030 y 700 millones en 2045 (2). De esos 463 millones, 4.2 millones fallecieron por DM, lo que representa el 11.3 % de las muertes por todas las causas (el equivalente a ocho muertes por minuto) y se estima que prácticamente la mitad de esas muertes (46.2 %, 1.9 millones) se produjeron en personas menores de 60 años (3).

El autocuidado es considerado uno de los factores más importantes para incrementar la calidad de vida de los pacientes con DM. Sin embargo, esto requiere que este tipo de pacientes adquieran los conocimientos y habilidades adecuadas para manejar su cuidado diario (4). Los pacientes con DM tienen grandes dificultades para realizar un correcto manejo dietético, tanto en calidad como en cantidad (5). La selección de comidas y las rutinas alimentarias están influenciadas por el conocimiento dietético de estos pacientes, lo que dificulta el automanejo de su dieta (6) y hace necesario un aumento de sus conocimientos con el fin de mejorar los resultados. En este sentido, las enfermeras juegan un papel crucial en la evaluación y seguimiento de los conocimientos dietéticos en esta población (7,8).

2. ANTECEDENTES

Existe un número limitado de instrumentos para evaluar el nivel de conocimiento de los pacientes con DM. Estos tienden a enfocarse en temas relacionados con la sintomatología de la enfermedad, el tratamiento farmacológico, el autocontrol de la glucemia, el ejercicio físico y la dieta (9). Por otro lado, uno de los principales factores que conducen a una mayor dificultad en el paciente con DM es el manejo dietético, debiendo ser necesario evaluar

previamente si el paciente tiene suficientes conocimientos sobre el tema para practicar una dieta saludable (10). Elegir el instrumento más apropiado para evaluar el nivel de conocimiento de los pacientes sobre su DM es un desafío, y la elección debe depender de las propiedades de medición y la viabilidad de los instrumentos (11).

Se realizó una búsqueda bibliográfica en diferentes bases de datos (Medline, Web of Science, Open Greay, Cochrane Library, base de datos COSMIN y Scielo) para encontrar una revisión sistemática sobre instrumentos para evaluar el conocimiento dietético en DM tipo 1 y 2 sin encontrar ningún resultado. En este contexto, no existe una revisión sistemática en la que se haya hecho una evaluación de todas las herramientas para evaluar del conocimiento dietético para la DM. Teniendo en cuenta este hecho, se efectuó una revisión para identificar, evaluar críticamente y resumir instrumentos para medir el conocimiento dietético en pacientes con DM tipo 1 o 2 basados en los estándares COSMIN más actualizados (Estándares basados en el COnsenso para la selección de INstrumentos de Medición de la salud) (12,13).

3. REVISIÓN SISTEMÁTICA

3.1 Objetivos

Esta revisión sistemática tuvo como objetivo identificar, evaluar críticamente y resumir instrumentos para medir el conocimiento dietético en pacientes con DM tipo 1 o 2.

3.2 Diseño

Se realizó una revisión sistemática de las propiedades de medición de los instrumentos que evalúan el nivel de conocimiento dietético en la población diabética tipo 1 y 2 de acuerdo con la lista más actualizada de verificación de cuatro puntos COSMIN para revisiones sistemáticas y los Elementos de Informes Preferidos para Estándares de Revisiones Sistemáticas y Metaanálisis (PRISMA) (12-14). El propósito de la revisión sistemática es localizar estudios existentes, seleccionar y evaluar contribuciones, analizar y sintetizar datos y reportar la

evidencia de tal manera que permita llegar a conclusiones razonablemente claras sobre lo que se sabe y lo que no.

La lista de verificación COSMIN se utilizó para evaluar las propiedades de medición de las medidas de resultado existentes informadas por el paciente para seleccionar el instrumento de medición de resultado más apropiado (12,13). El propósito de la declaración PRISMA es mejorar la presentación de informes sobre revisiones sistemáticas y metanálisis. Este protocolo ha sido inscrito en el Registro Prospectivo Internacional de Revisiones Sistemáticas (PROSPERO) (CRD42020186516).

3.3 Método de búsqueda

Se realizó una búsqueda exhaustiva durante marzo de 2020 en las diferentes bases de datos: Medline (a través de Pubmed), Web of Science, Open Grey y Scielo, por 2 revisores independientes. La estrategia de búsqueda incluía una combinación de los siguientes 5 aspectos: 1) búsqueda de población AND 2) búsqueda del constructo AND 3) búsqueda de instrumentos AND 4) filtro para las propiedades de medición AND 5) filtro de exclusión "NOT" (15). Con el fin de localizar el máximo número posible de herramientas y sus diferentes versiones, no se limitaron las fechas. Para realizar esta búsqueda se utilizaron los siguientes términos clave: "diabetes mellitus", "nutritional knowledge", "nutrition", "nutrition status", "dietary knowledge", "diet", "nutritional status", "questionnaire", "assessment tool", "instruments", "measure", "scale", "tool", "survey" and "test", adaptando esta estrategia a cada uno de los sistemas de búsqueda de las diferentes bases de datos. La búsqueda se limitó a artículos publicados en inglés y español. Además, se revisaron las listas de referencias de los artículos incluidos para identificar otros documentos relevantes (tabla 5).

Tabla 5. Estrategia de búsqueda.

Search strings for Pubmed

- 1 "Diabetes mellitus" [MeSH] OR diabetes [All Fields]
- 2 "nutritional knowledge" [All Fields] OR nutrition [All Fields] OR "nutrition status" [All Fields] OR "dietary knowledge" [All Fields] OR diet [All Fields]"nutritional status"[All Fields] OR "nutrition"[All Fields] OR "nutrition status"[All Fields] AND "questionnaire"[All Fields] AND "assessment tool"[All Fields] AND "diabetes"[All Fields] AND "humans"[MeSH Terms]
- 3 (Instrument[tiab] OR instruments[tiab] OR measure [tiab] OR measures [tiab] OR questionnaire[tiab] OR questionnaires[tiab] OR scale[tiab] OR scales[tiab] OR tool[tiab] OR tools[tiab] OR survey [tiab] OR test [tiab])
- 4 Instrumentation[sh] OR methods[sh] OR "Validation Studies"[pt] OR "Comparative Study"[pt] OR "psychometrics"[MeSH] OR psychometr*[tiab] OR clinimetr*[tw] OR clinometr*[tw] OR "outcome assessment (health care)"[MeSH] OR "outcome assessment"[tiab] OR "outcome measure*" [tw] OR "observer variation"[MeSH] OR "observer variation"[tiab] OR "Health Status Indicators"[Mesh] OR "reproducibility of results"[MeSH] OR reproducib*[tiab] OR "discriminant analysis"[MeSH] OR reliab*[tiab] OR unreliab*[tiab] OR valid*[tiab] OR coefficient[tiab] OR homogeneity[tiab] OR homogeneous[tiab] OR "internal consistency"[tiab] OR (cronbach*[tiab] AND (alpha[tiab] OR alphas[tiab])) OR (item[tiab] AND (correlation*[tiab] OR selection*[tiab] OR reduction*[tiab])) OR agreement[tiab] OR precision[tiab] OR imprecision[tiab] OR "precise values"[tiab] OR test-retest[tiab] OR (test[tiab] AND retest[tiab]) OR (reliab* [tiab] AND (test[tiab] OR retest[tiab])) OR stability[tiab] OR interrater[tiab] OR inter-rater[tiab] OR intrarater[tiab] OR intra-rater[tiab] OR intertester[tiab] OR inter-tester[tiab] OR intratester[tiab] OR intra-tester[tiab] OR interobserver[tiab] OR inter-observer[tiab] OR intraobserver[tiab] OR intraobserver[tiab] OR intertechnician[tiab] OR inter-technician[tiab] OR intratechnician[tiab] OR intra-technician[tiab] OR interexaminer[tiab] OR inter-examiner[tiab] OR intraexaminer[tiab] OR intra-examiner[tiab] OR interassay[tiab] OR inter-assay[tiab] OR intraassay[tiab] OR intra-assay[tiab] OR interindividual[tiab] OR inter-individual[tiab] OR intraindividual[tiab] OR intra-individual[tiab] OR interparticipant[tiab] OR inter-participant[tiab] OR intraparticipant[tiab] OR intra-participant[tiab] OR kappa[tiab] OR kappa's[tiab] OR kappas[tiab] OR repeatab*[tiab] OR ((replicab*[tiab] OR repeated[tiab]) AND (measure[tiab] OR measures[tiab] OR findings[tiab] OR result[tiab] OR results[tiab] OR test[tiab] OR tests[tiab])) OR generaliza*[tiab] OR generalisa*[tiab] OR concordance[tiab] OR (intraclass[tiab] AND correlation*[tiab]) OR discriminative[tiab] OR "known group"[tiab] OR "factor analysis"[tiab] OR "factor analyses"[tiab] OR dimension*[tiab] OR subscale*[tiab] OR (multitrait[tiab] AND scaling[tiab] AND (analysis[tiab] OR analyses[tiab])) OR "item discriminant"[tiab] OR "interscale correlation*" [tiab] OR error[tiab] OR errors[tiab] OR "individual variability"[tiab] OR (variability[tiab] AND (analysis[tiab] OR values[tiab])) OR (uncertainty[tiab] AND (measurement[tiab] OR measuring[tiab])) OR "standard error of measurement"[tiab] OR sensitiv*[tiab] OR responsive*[tiab] OR ((minimal[tiab] OR minimally[tiab] OR clinical[tiab] OR clinically[tiab]) AND (important[tiab] OR significant[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR (small*[tiab] AND (real[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR "meaningful change" [tiab] OR "ceiling effect"[tiab] OR "floor effect"[tiab] OR "Item response model"[tiab] OR IRT[tiab] OR Rasch[tiab] OR "Differential item functioning"[tiab] OR DIF[tiab] OR "computer adaptive testing"[tiab] OR "item bank"[tiab]
- 5 #1 AND #2 AND #3 AND #4
- 6 "Protocol"[Publication Type] OR "addresses"[Publication Type] OR "biography"[Publication Type] OR "case reports"[Publication Type] OR "comment"[Publication Type] OR "editorial"[Publication Type] OR "congresses"[Publication Type] OR "consensus development conference"[Publication Type] OR "consensus development conference"[Publication Type] OR "practice guideline"[Publication Type]) OR "suffering from"[tiab] OR "animals"[MeSH]
- 7 #5 NOT #6
- 8 FILTER: Language (English and Spanish)

Search strings for Web of Science (WoS)

- 1 TI=("Diabetes Mellitus" OR "diabetes")
- 2 TS=(nutritional knowledge OR nutrition OR nutrition status OR dietary knowledge OR diet)
- 3 TS=(Instrument OR instruments OR measure OR measures OR questionnaire OR questionnaires OR scale OR scales OR tool OR tools OR survey OR test)
- 4 TS=("Instrumentation" OR "methods" OR "Validation Studies" OR "Comparative Study" OR "psychometrics" OR psychometr* OR clinimetr* OR clinometr* OR "outcome assessment (health care)" OR "outcome assessment" OR outcome measure* OR "observer variation" OR "observer variation" OR "Health Status Indicators" OR "reproducibility of results" OR reproducib* OR "discriminant analysis" OR reliab* OR unreliab* OR valid* OR "coefficient" OR "homogeneity" OR "homogeneous" OR "internal consistency" OR cronbach* AND "alpha" OR "alphas" OR "item" AND (correlation OR selection OR reduction) OR "agreement" OR "precision" OR "imprecision" OR "precise values" OR "test-retest" OR (test AND retest OR (reliab AND (test OR retest)) OR "stability" OR "interrater" OR "inter-rater" OR "intrarater" OR "intra-rater" OR "intertester" OR "inter-tester" OR "intratester" OR "intra-tester" OR "interobserver" OR "inter-observer" OR "intraobserver" OR "intraobserver" OR "intertechician" OR "inter-technician" OR "intratechnician" OR "intra-technician" OR "interexaminer" OR "inter-examiner" OR "intraexaminer" OR "intra-examiner" OR "interassay" OR "inter-assay" OR "intraassay" OR "intra-assay" OR "interindividual" OR "inter-individual" OR "intraindividual" OR "intra-individual" OR "interparticipant" OR "inter-participant" OR "intraparticipant" OR "intra-participant" OR "kappa" OR "kappa's" OR "kappas" OR repeatab* OR ((replicab OR repeated) AND (measure OR measures OR findings OR result OR results OR test OR tests)) OR generaliza* OR generalisa* OR concordance* OR (intraclass AND correlation)OR "discriminative" OR "known group" OR "factor analysis" OR "factor analyses" OR dimension* OR "subscale" OR (multitrait AND scaling AND (analysis OR analyses)) OR "item discriminant" OR interscale correlation* OR "error" OR "errors" OR "individual variability" OR (variability AND (analysis OR values)) OR (uncertainty AND (measurement OR measuring)) OR "standard error of measurement" OR sensitiv* OR responsive* OR ((minimal OR minimally OR clinical OR clinically) AND (important OR significant OR detectable) AND (change OR difference)) OR (small* AND (real OR detectable) AND (change OR difference)) OR "meaningful change" OR "ceiling effect" OR "floor effect" OR "Item response model" OR IRT OR Rasch OR "Differential item functioning" OR DIF OR "computer adaptive testing" OR "item bank" OR "cross-cultural equivalence")
- 5 #1 AND #2 AND #3 AND #4
- 6 TI=("Protocol" OR "addresses" OR "biography" OR "case reports" OR "comment" OR "editorial" OR "congresses" OR "consensus development conference" OR "consensus development conference" OR "practice guideline")
- 7 #5 NOT #6
- 8 FILTER: Language (English and Spanish)

Search strings for Open Grey

- 1 ("Diabetes Mellitus" OR "diabetes")
- 2 (nutritional knowledge OR nutrition OR nutrition status OR dietary knowledge OR diet)
- 3 (Instrument OR instruments OR measure OR measures OR questionnaire OR questionnaires OR scale OR scales OR tool OR tools OR survey OR test)
- 4 ("Instrumentation" OR "methods" OR "Validation Studies" OR "Comparative Study" OR "psychometrics" OR psychometr* OR clinimetr* OR clinometr* OR "outcome assessment (health care)" OR "outcome assessment" OR outcome measure* OR "observer variation" OR "observer variation" OR "Health Status Indicators" OR "reproducibility of results" OR reproducib* OR "discriminant analysis" OR reliab* OR unreliab* OR valid* OR "coefficient" OR "homogeneity" OR "homogeneous" OR "internal consistency" OR cronbach* AND "alpha" OR "alphas" OR "item" AND (correlation OR selection OR reduction) OR "agreement" OR "precision" OR "imprecision" OR "precise values" OR "test-retest" OR (test AND retest OR (reliab AND (test OR retest)) OR "stability" OR "interrater" OR "inter-rater" OR "intrarater" OR "intra-rater" OR "intertester" OR "inter-tester" OR "intratester" OR "intra-tester" OR "interobserver" OR "inter-observer" OR "intraobserver" OR "intraobserver" OR "intertechician" OR "inter-technician" OR "intratechnician" OR "intra-technician" OR "interexaminer" OR "inter-examiner" OR "intraexaminer" OR "intra-examiner" OR "interassay" OR "inter-assay" OR "intraassay" OR "intra-assay" OR "interindividual" OR "inter-individual" OR "intraindividual" OR "intra-individual" OR "interparticipant" OR "inter-participant" OR "intraparticipant" OR "intra-participant" OR "kappa" OR "kappa's" OR "kappas" OR repeatab* OR ((replicab OR repeated) AND (measure OR measures OR findings OR result OR results OR test OR tests)) OR generaliza* OR generalisa* OR concordance* OR (intra OR inter) AND correlation) OR "discriminative" OR "known group" OR "factor analysis" OR "factor analyses" OR dimension* OR "subscale" OR (multitrait AND scaling AND (analysis OR analyses)) OR "item discriminant" OR interscale correlation* OR "error" OR "errors" OR "individual variability" OR (variability AND (analysis OR values)) OR (uncertainty AND (measurement OR measuring)) OR "standard error of measurement" OR sensitiv* OR responsive* OR ((minimal OR minimally OR clinical OR clinically) AND (important OR significant OR detectable) AND (change OR difference)) OR (small* AND (real OR detectable) AND (change OR difference)) OR "meaningful change" OR "ceiling effect" OR "floor effect" OR "Item response model" OR IRT OR Rasch OR "Differential item functioning" OR DIF OR "computer adaptive testing" OR "item bank" OR "cross-cultural equivalence")
- 5 #1 AND #2 AND #3 AND #4
- 6 ("Protocol" OR "addresses" OR "biography" OR "case reports" OR "comment" OR "editorial" OR "congresses" OR "consensus development conference" OR "consensus development conference" OR "practice guideline")
- 7 #5 NOT #6
- 8 FILTER: Language (English and Spanish)

Search strings for Scielo

- 1 ("Diabetes Mellitus" OR "diabetes")
- 2 (conocimiento nutricional OR nutricion OR estado nutricional OR conocimiento dietetico OR dieta)
- 3 (Instrumento OR instrumentos OR medida OR medidas OR cuestionario OR cuestionarios OR escala OR escalas OR herramienta OR herramientas OR encuesta OR test)
- 4 ("Instrumentacion" OR "metodos" OR "estudios de validacion" OR "estudios comparativos" OR "psicometria" OR psicometr* OR clinimetr* OR clinometr* OR "resultados de valoracion (cuidado de salud)" OR "resultados de valoracion" OR "variables de resultado" OR "variacion del observador" OR "indicadores de estado de salud" OR "reproductibilidad de los resultados" OR reproducib* OR "analisis discriminativo" OR fiab* OR valid* OR "coeficiente" OR "homogeneidad" OR "consistencia interna" OR cronbach* AND "alpha" OR "alphas" OR "item" AND (correlacion OR seleccion OR reduccion) OR "acuerdo" OR "precision" OR "imprecision" OR "test-retest" OR (test AND retest OR (fiab* AND (test OR retest))) OR "estabilidad" OR "interevaluador" OR "inter-evaluador" OR "intraevaluador" OR "intra-evaluador" OR "intertester" OR "inter-tester" OR "intratester" OR "intra-tester" OR "interobservador" OR "inter-observador" OR "intraobservador" OR "intraobservador" OR "intertecnico" OR "inter-tecnico" OR "intratecnico" OR "intra-tecnico" OR "interexaminador" OR "inter-examinador" OR "intraexaminador" OR "intra-examinador" OR "interensayo" OR "inter-ensayo" OR "intraensayo" OR "intra-ensayo" OR "interindividual" OR "inter-individual" OR "intraindividual" OR "intra-individual" OR "interparticipante" OR "inter-participante" OR "intraparticipante" OR "intra-participante" OR "kappa" OR "kappa's" OR "kappas" OR repetib* OR ((replicab OR repetibilidad) AND (medida OR medidas OR resultados OR resultado OR test OR tests)) OR generaliza* OR generalisa* OR concordancia* OR (intraclase AND correlacion OR "discriminativa" OR "factor analisis" OR dimension* OR "subescala" OR (multi-retrato AND escalada AND analisis) OR "item discriminante" OR correlacion interescalas* OR "error" OR "errores" OR "variabilidad individual" OR (variabilidad AND (analisis OR valores)) OR (incertidumbre AND (medicion OR medida)) OR "error standard de medida" OR sensib* OR responsive* OR ((minimal OR clinico OR clinical) AND (importante OR significativo OR detectable) AND (cambio OR diferencia)) OR (pequeño* AND (real OR detectable) AND (cambio OR diferencia)) OR "cambio significativo" OR "efecto techo" OR "efecto suelo" OR "modelo de respuesta" OR IRT OR Rasch OR "Funcionamiento diferencial de elementos" OR DIF OR "pruebas adaptativas por computadora" OR "banco de articulos" OR "cross-cultural equivalencia")
- 5 #1 AND #2 AND #3 AND #4
- 6 ("Protocolo" OR "direcciones" OR "biografia" OR "estudios de caso" OR "commentarios" OR "editorial" OR "congresos" OR "conferencia" OR "guía de práctica clínica")
- 7 #5 NOT #6
- 8 FILTER: Language (English and Spanish)

3.4 Criterio de elegibilidad

Los criterios de inclusión para este estudio fueron:

- Tipo de participantes: pacientes con DM tipo 1 o 2.
- Tipo de resultados: nivel de conocimiento dietético en DM tipo 1 o 2.
- Tipo de estudios: estudios que habían descrito el desarrollo de instrumentos que miden el nivel de conocimiento dietético en DM tipo 1 o 2 y evaluaron al menos una propiedad psicométrica de la lista de verificación COSMIN.
- Tipo de instrumento: instrumentos que evalúan el nivel de conocimiento dietético en DM tipo 1 o 2.

Los criterios de exclusión para este estudio fueron:

- Instrumento no disponible en inglés o español.
- Editoriales, biografías, discursos, comentarios y conferencias, informes de casos, guías de práctica, tesis doctorales, artículos de opinión o artículos en los que no se haya utilizado metodología selectiva.

Los títulos y los resúmenes fueron seleccionados de forma independiente por 2 revisores. El artículo fue recuperado en su totalidad cuando no fue posible tomar decisiones solo sobre el resumen y el título. Las discrepancias se resolvieron mediante discusión con un tercer revisor.

3.5 Resultados de la búsqueda

Se identificaron un total de 2065 artículos, de los cuales se examinaron 1433 títulos y resúmenes. De estos, 1409 artículos fueron excluidos tras esta evaluación. Posteriormente, se evaluaron 24 artículos a texto completo. De estos, 2 artículos fueron excluidos después de revisar el texto completo (6,16). Finalmente, 22 estudios fueron incluidos en esta revisión sistemática. El diagrama de flujo se muestra en la figura 16.

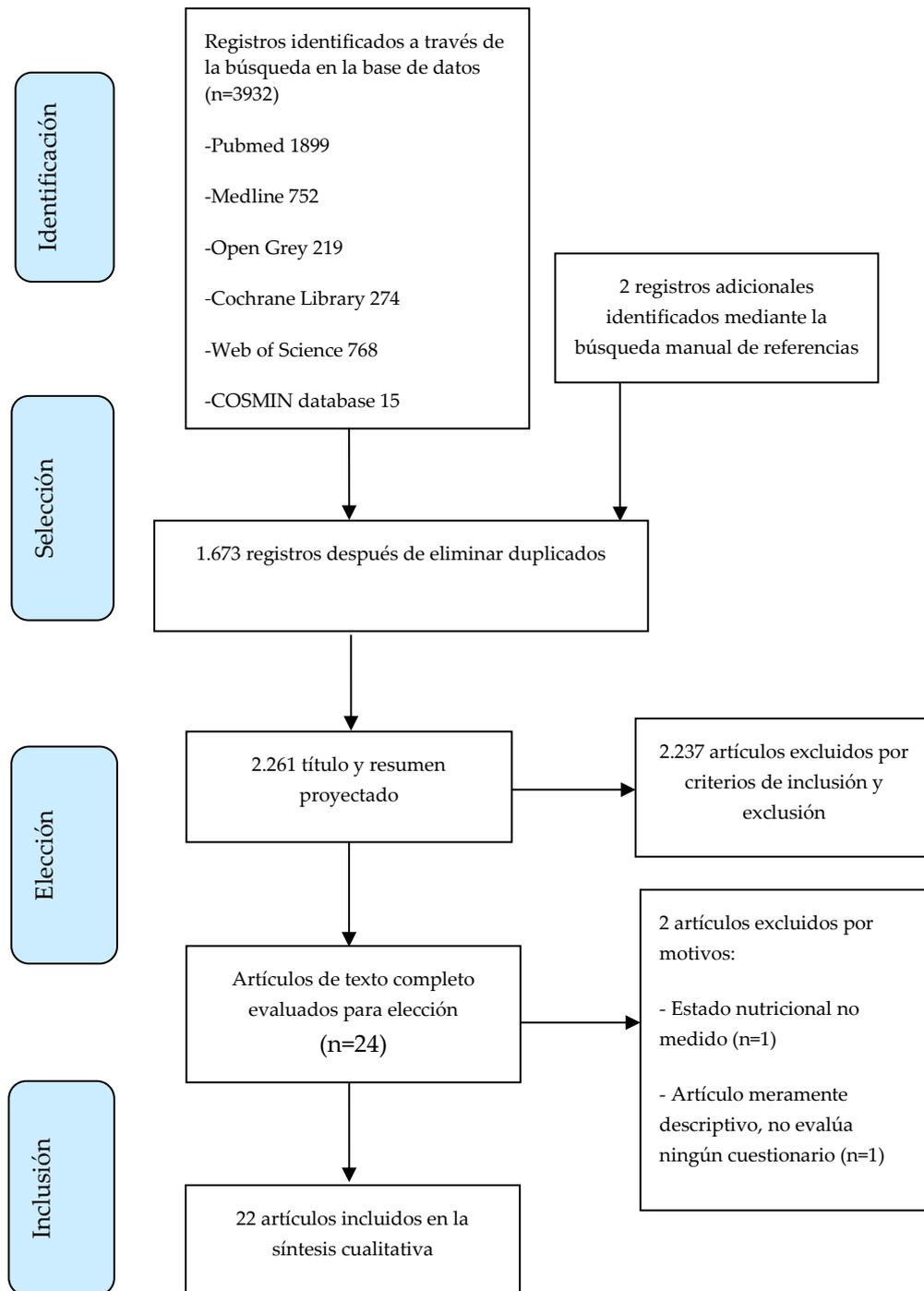


Figura 16. Diagrama de flujo de la selección del estudio.

3.6 Evaluación de la calidad

Los estudios incluidos en la revisión fueron evaluados por Pedro Montagut Martínez, David Pérez Cruzado y Daniel Gutiérrez Sánchez, de acuerdo con la lista más actualizada de verificación de cuatro puntos COSMIN para revisiones sistemáticas (disponible en <https://www.cosmin.nl/>) (13). Esta lista contiene 9 cajas con 5 - 18 elementos y cada caja ofrece una puntuación de calidad metodológica para validez de contenido, validez estructural, consistencia interna, validez transcultural, error de medición, fiabilidad, validez de criterio, prueba de hipótesis y sensibilidad al cambio. El contenido evaluado en cada casilla se puede calificar como “muy bueno”, “bueno”, “dudoso” e “inadecuado”. La puntuación de cada propiedad psicométrica se obtiene tomando la calificación más baja de cualquier ítem para cada propiedad (13). Se rechazó cualquier artículo según la calidad metodológica del instrumento incluido.

3.7 Extracción de datos

Se revisaron todos los títulos y resúmenes de los artículos identificados en esta revisión. Los artículos a texto completo fueron revisados de forma independiente por 3 autores y las discrepancias se resolvieron por consenso. La información de cada artículo se extrajo basándonos en los siguientes aspectos: nombre de la herramienta utilizada, población de estudio, características de la herramienta, análisis de propiedades psicométricas, puntuación COSMIN y resultados estadísticos de la investigación.

3.8 Síntesis

Los datos sobre las propiedades psicométricas de los artículos que cumplieron con los criterios de inclusión se calificaron de acuerdo con la lista de verificación COSMIN (13). Las puntuaciones de cada una de las propiedades psicométricas de cada artículo fueron establecidas por 3 revisores independientes y las diferencias fueron resueltas mediante discusión y consenso. Posteriormente, se calificó el resultado de cada una de las propiedades de medición de cada estudio según los criterios de Terwee (11), clasificando cada uno de los resultados como “suficiente” (+), “insuficiente” (-) o “indeterminado” (?). Finalmente,

después de determinar la calidad metodológica de cada uno de los estudios individuales y evaluar las propiedades de medición, los resultados se resumieron cualitativamente para determinar tanto la calidad de la evidencia como la fuerza de las recomendaciones mediante la calificación de “alta”, “moderada”, “baja” o “muy baja” según el riesgo de sesgo, inconsistencia, imprecisión y ausencia de evidencia directa utilizando el método GRADE (17). Las puntuaciones para cada una de las propiedades psicométricas de cada artículo fueron asignadas por 3 revisores independientes, y las diferencias se resolvieron mediante discusión y consenso.

4. RESULTADOS

4.1 Instrumentos identificados para medir el conocimiento dietético

Se identificaron 9 instrumentos que fueron diseñados específicamente para evaluar el conocimiento dietético en pacientes con DM tipo 1 o 2: Diabetes Patient Knowledge Test (DPKT) (18), Diabetes Knowledge Assessment Scale (DKN) (19), Michigan Diabetes Knowledge Test (DKT) (1,20–29), Diabetes Knowledge Questionnaire (DKQ) (30–33), Diabetes Knowledge Assessment Test (DKAT) (10), Personal Diabetes Questionnaire (PDQ) (34,35), Diabetes Mellitus Knowledge (DMK) (33), Conocidiet-Diabetes (9) y Diabetes-related Nutrition Knowledge (DRNK) (36).

4.2 Características de los instrumentos incluidos para medir el conocimiento dietético

Las características de los instrumentos incluidos para medir el conocimiento dietético y los estudios se presentan en la tabla 6.

En la mayoría de los casos el idioma original utilizado fue el inglés, aunque encontramos varios de estos cuestionarios en los que se llevó a cabo una validación transcultural: versiones en islandés (21), malayo (22), árabe (1, 25, 28) y turco (29) de DKT, una versión en español (30) y urdu (32) de DKQ, y una versión china (34) de PDQ. Se realizaron diferentes versiones de DKT (1,20–29), DKQ (30–33) y PDQ (34,35), cada una de las cuales variaba en el número de ítems. La

mayoría de estos estudios se efectuaron en el ámbito hospitalario, comunitario y de atención primaria en pacientes con DM tipo 1 (21,27), DM tipo 2 (22,24,25,30,32–34,36) o DM tipo 1 y 2 (1,9,10,18–20,26,28,29,35).

La mayoría de los estudios se realizaron en Asia, donde DKT se utilizó en mayor número de ocasiones: en Malasia (22), Arabia Saudita (1,25), India (27), Emiratos Árabes Unidos (28) y Turquía (29). Lo mismo sucedió en el caso de DKQ, que se empleó para 3 estudios: en Malasia (31), Pakistán (32) y Arabia Saudita (33). El resto de cuestionarios usados fueron PDQ en China (34), DMK en Arabia Saudita (33) y finalmente DRNK en Singapur (36). También se utilizó DKT, específicamente en Estados Unidos (20,26), donde se desarrollaron otros estudios con las herramientas DPKT (18), DKAT (10) y PDQ (35). Solo uno de los estudios se realizó fuera de los Estados Unidos, concretamente en México, donde se empleó DKQ (30). Además de los países citados con anterioridad, DKT se usó en Europa, esta vez en Islandia (21) y Reino Unido (23). Otro de los instrumentos utilizados, excepcionalmente en España, fue el Conocidiet-Diabetes (9). Finalmente, 2 estudios efectuados en Australia incluyeron el DKN (19) y DKT (24).

En cuanto al tamaño muestral de los estudios, existió gran heterogeneidad: estos oscilaron entre los 10 participantes incluidos en una de las versiones de DKT (28) y los 956 pacientes del estudio de DKN (19).

Con relación al número de ítems, los cuestionarios variaron desde los dos más breves, constituidos por los 14 ítems de las versiones más cortas de DKT, hasta los 89 ítems de DKQ, el más amplio (31). Por otro lado, la versión más larga de DKT presentó 34 ítems, mientras que la versión más corta de DKQ tuvo 20 ítems (33). Finalmente, las dos versiones de PDQ tenían 45 (34) y 49 ítems (35).

Por el tipo de pregunta/respuesta utilizada en cada uno de los cuestionarios, la mayoría de ítems fueron preguntas con respuestas de opción múltiple, incluido DPKT (18), DKN (19), DKT (1,20–22,24–27,29), Revised Diabetes Knowledge Scale (RDKS) (23) (una versión corta de DKT), DRNK (36) y DKQ (33). El formato en el resto de DKQ (30–32) eran preguntas cerradas que tenían tres posibles opciones de respuesta: 1) Sí, 2) No y 3) "No sé". Para PDQ (34,35) el tipo de respuesta se dio mediante una escala tipo Likert de 6 puntos, con un rango de 1 (Nunca) a 6 (Una o más veces al día). En el resto de los cuestionarios, el tipo de respuesta fue verdadero/falso para dos versiones cortas de DKT (Simplified Diabetes

Knowledge Scale (SDKS) (23) y Simplified Diabetes Knowledge Test (SDKT) (28)), DKAT (10) y Conocidiet-Diabetes (9). Finalmente, y a diferencia del resto de cuestionarios, DMK (33) tenía como respuesta tres posibles opciones codificadas, 1 = "Sí", 0 = "No", y "No sé".

Tabla 6. Características de los estudios y descripción de los instrumentos de conocimiento de la diabetes.

Estudio (autor y año)	Poplación / tipo de DM	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Diabetes Patient Knowledge Test (DPKT)						
Hess et al. (1983)	956 mediciones a pacientes con diabetes tipo 1 y 2	Estados Unidos: pacientes hospitalizados en el Hospital Universitario	La prueba de conocimiento del paciente con diabetes (DPKT) se compone de cinco subgrupos: "Carbohidratos" (seis elementos relacionados con la identificación de alimentos con alto contenido de carbohidratos y el manejo de carbohidratos en los días de enfermedad), "azúcar en la sangre" (contiene 10 elementos relacionados con signos, causas y manejo del estado glucémico anormal), "conceptos básicos" (incluye 11 ítems básicos para objetivos de tratamiento, análisis de orina, complicaciones y estado de control), "intercambio de alimentos" (6 ítems) y "administración de insulina" (5 ítems sobre manejo y administración de insulina), en total 38 ítems.	1-Validez de contenido 2-Validez estructural 3-Consistencia interna 4-Prueba de hipótesis 5-Sensibilidad al cambio	1-Inadecuado 2-Inadecuado 3-Inadecuado 4-Dudoso 5-Dudoso	- Un análisis factorial del eje principal de 210 casos dio como resultado una solución de 5 factores que representa el 37% de la varianza en la respuesta de la prueba. - α Cronbach para ambas versiones = 0.89. - Las correlaciones entre la puntuación de la prueba y el nivel educativo oscilaron entre 0.07 para el factor de intercambio y 0.20 para el factor Básico. Las correlaciones de esta baja magnitud indicaron que la prueba midió más que el nivel de educación general. Además, las correlaciones más bajas se asociaron con un conocimiento más especializado y las correlaciones más altas con el conocimiento de aspectos más generales de la diabetes.

Estudio (autor y año)	Poplación / tipo de DM	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Diabetes Knowledge Assessment Scale (DKN)						
Dunn et al. (1984)	300 pacientes con diabetes tipo 1 y 2	Australia: Los pacientes fueron reclutados en el Royal Prince Alfred Hospital	El DKN-A es un cuestionario autoadministrado con 15 ítems. Presenta cinco categorías amplias: fisiología básica, hipoglucemia, grupos de alimentos y sus sustituciones, manejo de la diabetes en caso de que ocurra alguna otra enfermedad y principios generales del cuidado de la enfermedad. La escala de medición es de 0 a 15. A cada ítem se le puede asignar una puntuación de 1 para la respuesta correcta y 0 para la incorrecta. Los ítems 1-12 requieren una sola respuesta correcta. Para los ítems del 13 al 15, solo algunas respuestas son correctas y todas deben conferirse para obtener la puntuación 1. Una puntuación superior a ocho indica conocimiento sobre diabetes.	1-Validez de contenido 2-Consistencia interna 3- Prueba de hipótesis 4-Sensibilidad al cambio	1-Adecuado 2-Inadecuado 3- Dudoso 4- Dudoso	- α Cronbach = 0.92. - Las formas paralelas DKNA, DKNB y DKNC, para cada uno de los 15 elementos seleccionados del conjunto principal, tenían coeficientes alfa superiores a 0.82 y se correlacionaban 0.90 entre sí.

Estudio (autor y año)	Poplación / tipo de DM	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Michigan Diabetes Knowledge Test (DKT)						
Fitzgerald et al. (1998)	811 pacientes con diabetes tipo 1 y 2 (312 comunitarios y 499 MDPH)	Estados Unidos: Los pacientes fueron reclutados de la comunidad y del Departamento de Salud Pública de Michigan.	La prueba de conocimiento sobre diabetes consta de 23 elementos. Los primeros 14 elementos se aplican a todos los pacientes y los 9 elementos restantes son relevantes para aquellos que usan insulina. A cada ítem se le asigna una puntuación de 1 para una respuesta correcta y 0 para una respuesta incorrecta. El conocimiento de la diabetes se califica sumando el número de respuestas correctas y convirtiendo el puntaje bruto en porcentaje correcto, 23 elementos correctos dan respuestas 100% correctas. La prueba de 23 ítems tarda aproximadamente 15 minutos en completarse. En pacientes que no toman insulina se aplicaría una versión de 14 ítems según el autor.	1-Consistencia interna 2-Prueba de hipótesis 3-Sensibilidad al cambio	1-Inadecuado 2-Adecuado 3-Dudoso	- α Cronbach para la prueba general para la comunidad, MDPH y total fue 0.70, 0.71 y 0.71, respectivamente, y para los usuarios de insulina fue 0.74, 0.76 y 0.75, respectivamente. - En la muestra comunitaria, los pacientes con DM 1 obtuvieron puntuaciones más altas que los pacientes con diabetes tipo 2 en la prueba general y la subescala de uso de insulina. En la muestra del departamento de salud, los pacientes con tipo 1 obtuvieron puntuaciones más altas que los pacientes con tipo 2 en la subescala de uso de insulina. Para ambas muestras, los puntajes aumentaron a medida que aumentaron los años de educación formal completados, y los pacientes que recibieron educación sobre diabetes obtuvieron puntajes más altos que los pacientes que no la recibieron.

Estudio (autor y año)	Poplación / tipo de DM	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Michigan Diabetes Knowledge Test (DKT)						
Sigurdardottir et al. (2008)	Personas con diabetes insulino dependientes	Islandia: clínicas de diabetes.	Como anteriormente	1-Validez de contenido 2-Validez estructural 3-Consistencia interna 4- Validez transcultural / invariancia de medición 5-Prueba de hipótesis 6-Sensibilidad al cambio	1-Inadecuado 2-Inadecuado 3-Inadecuado 4- Dudoso 5- Dudoso 6- Dudoso	- El análisis factorial reveló 2 factores: (a) angustia en relación con la vida con diabetes y (b) angustia en relación con el manejo de la diabetes. - α Cronbach = 0.94. - El control metabólico se correlacionó significativamente con las puntuaciones de PAID ($r = 0.50$, $p < 0.001$, IC: 0.33 a 0.76). La puntuación de todos los elementos en PAID se correlacionó significativamente con el nivel de HbA1c. El control de la educación, el estado de vida y la puntuación media de la subescala psicosocial del DES demostró casi la misma correlación entre las puntuaciones de PAID y el nivel de HbA1c ($r = 0.49$, $p < 0.001$).

Estudio (autor y año)	Poplación / tipo de DM	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Michigan Diabetes Knowledge Test (DKT)						
Al-Qazaz et al. (2010)	307 pacientes con diabetes tipo 2	Malasia: Clínica de Diabetes del Hospital General de Penang. Pacientes ambulatorios.	Versión malaya de la prueba MDKT	1-Validez de contenido 2-Consistencia interna 3- Validez transcultural / invariancia de medición 4-Fiabilidad 5- Prueba de hipótesis 6-Sensibilidad al cambio	1-Dudoso 2-Inadecuado 3-Dudoso 4-Inadecuado 5-Dudoso 6-Dudoso	- α Cronbach = 0.702 - CCI = 0.894 (p < 0.001) - Se encontró una relación significativa entre las categorías MDKT y las categorías HbA1c (chi-cuadrado = 21.626; p \geq 0.001).

Estudio (autor y año)	Poplación / tipo de DM	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Michigan Diabetes Knowledge Test (DKT)						
Collins et al. (2010)	99 pacientes con diabetes tipo 1 y 2	Reino Unido: muestra de conveniencia de pacientes que asistieron al Birmingham Clínica ambulatoria del Heartlands Hospital	La Escala Revisada de Conocimiento de la Diabetes (RDKS) consta de 17 preguntas de opción múltiple (elementos básicos) para evaluar el conocimiento del paciente relacionado con la diabetes, incluidos elementos sobre dieta, control de glucosa en sangre, ejercicio, toma de medicamentos y complicaciones, con tres elementos adicionales para pacientes que toman insulina. Cada pregunta contiene cuatro respuestas posibles para que el encuestado elija. Esta escala es una versión de la Escala de conocimiento de la diabetes de Michigan original (MDKS) o Prueba de conocimiento de la diabetes (DKT), llamada así hasta 2016. Más tarde pasó a llamarse DKT2.	1-Validez estructural 2-Consistencia interna 3-Prueba de hipótesis 4-Sensibilidad al cambio	1-Dudoso 2-Inadecuado 3- Inadecuado 4- Inadecuado	- El análisis de Rasch indicó que los ítems 1, 13, 19 y 9 fueron los más fáciles de responder, mientras que los ítems 5, 2, 6, 20 fueron los más difíciles de responder, lo que concuerda estrechamente con el porcentaje de respuestas correctas. La bondad de ajuste, determinada por 200 bootstraps de la estadística de la prueba χ^2 de Pearson, no indicó preocupaciones por la falta de ajuste ($P = 0.37$). - α Cronbach = 0.61 - Las correlaciones de los ítems con la puntuación total de la prueba de conocimientos variaron entre 0.23 y 0.45 para la Escala de conocimientos sobre diabetes revisada. Las correlaciones totales de los ítems corregidos (con el ítem eliminado) fueron > 0.2 , excepto para los ítems 7, 8, 12, 15 y 16.

Estudio (autor y año)	Poplación / tipo de DM	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Michigan Diabetes Knowledge Test (DKT)						
Collins et al. (2010)	99 pacientes con diabetes tipo 1 y 2	Reino Unido: muestra de conveniencia de pacientes que asistieron al Birmingham Clínica ambulatoria de diabetes del Heartlands Hospital	Se elabora una Escala Simplificada de Conocimiento sobre la Diabetes (SDKS) en la que las respuestas de opción múltiple de la Escala de Conocimiento sobre la Diabetes revisada se reemplazan por un formato "verdadero o falso", con la excepción del ítem 17 ("Se da cuenta justo antes del almuerzo que olvidó tomar su insulina antes del desayuno. ¿Qué debe hacer ahora? "), que no pudo responderse en un formato de verdadero / falso y, por lo tanto, está excluida.	1-Validez estructural 2-Consistencia interna 3-Prueba de hipótesis 4-Sensibilidad al cambio	1-Dudoso 2-Inadecuado 3- Inadecuado 4- Inadecuado	- El análisis de Rasch de la Escala Simplificada de Conocimiento sobre Diabetes indicó que los ítems 1, 13, 19 y 9 fueron los más fáciles de responder, mientras que los ítems 5, 2, 6, 20 fueron los más difíciles de responder, lo que concuerda estrechamente con el porcentaje de respuestas correctas. La bondad de ajuste, determinada por 200 bootstraps de la estadística de la prueba χ^2 de Pearson, no indicó preocupaciones por la falta de ajuste ($P = 0.37$). - α Cronbach = 0.61 - Las correlaciones de los ítems con la puntuación total de la prueba de conocimientos variaron de 0.26 a 0.58 en la Escala de conocimientos sobre diabetes simplificada. Las correlaciones totales de los ítems corregidos fueron > 0.2 , excepto para los ítems 7, 8 y 20.

Estudio (autor y año)	Poplación / tipo de DM	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Michigan Diabetes Knowledge Test (DKT)						
Fenwick et al. (2013)	181 pacientes con diabetes tipo 2	Australia: Los participantes fueron reclutados en clínicas oftalmológicas especializadas en el Royal Victorian Eye and Ear Hospital (RVEEH) y Australia Victoria.	La prueba de conocimiento sobre diabetes consta de 23 elementos. Los primeros 14 elementos se aplican a todos los pacientes y los 9 elementos restantes son relevantes para aquellos que usan insulina. A cada ítem se le asigna una puntuación de 1 para una respuesta correcta y 0 para una respuesta incorrecta. El conocimiento de la diabetes se califica sumando el número de respuestas correctas y convirtiendo el puntaje bruto en porcentaje correcto, 23 elementos correctos dan respuestas 100% correctas. La prueba de 23 ítems tarda aproximadamente 15 minutos en completarse.	1-Validez estructural 2-Prueba de hipótesis al cambio 3-Sensibilidad	1-Dudoso 2-Dudoso 3-Dudoso	- La precisión de la escala fue adecuada (PSI = 2.01 y PR = 0.80), y no hubo ítems desajustados ni evidencia de multidimensionalidad (varianza bruta explicada por las medidas 50.5% y varianza inexplicable en los valores propios del primer contraste 1.6). Es una escala de un solo factor. - Un nivel más alto de HbA1c se asoció de forma independiente con un peor conocimiento de la diabetes. Ser revisado por un oftalmólogo y la membresía de NDSS no se asociaron de forma independiente con el conocimiento de la diabetes. Los factores sociodemográficos, clínicos y de uso de servicios se asociaron de forma independiente con el conocimiento de la diabetes.

Estudio (autor y año)	Poplación / tipo de DM	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Michigan Diabetes Knowledge Test (DKT)						
Alhaiti et al. (2016)	139 participantes con diabetes tipo 2	Arabia Saudita: Participantes reclutados a través de King Fahad Medical City en Riyadh.	El DKT es un instrumento de 23 ítems diseñado para evaluar el conocimiento del paciente sobre la diabetes en relación con la dieta, el ejercicio, los niveles de glucosa en sangre y las pruebas y las actividades de autocuidado. Los primeros 14 elementos se aplican a todos los pacientes y los 9 elementos restantes son relevantes para aquellos que usan insulina. Las pruebas de 23 ítems son opciones de opción múltiple y se necesitan aproximadamente 15 minutos para completarlas.	1-Validez de contenido 2-Consistencia interna 3- Validez transcultural / invariancia de medición 4- Fiabilidad	1-Dudoso 2-Inadecuado 3-Dudoso 4- Inadecuado	- α Cronbach= 0.75 - ICC=0.90
Al Qahtani et al. (2016)	90 pacientes con diabetes tipo 1 y 2	Arabia Saudita: los pacientes fueron entrevistados en un centro de atención primaria	Versión árabe de las pruebas MDKT	1-Consistencia interna 2- Validez transcultural / invariancia de medición	1-Inadecuado 2-Inadecuado	- α Cronbach= 0.60.

Estudio (autor y año)	Poplación / tipo de DM	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Michigan Diabetes Knowledge Test (DKT)						
Fitzgerald et al. (2016)	190 pacientes con diabetes tipo 1 y 2	Estados Unidos: Se utilizaron dos muestras para evaluar. (Qualtrics LCC) (n = 101) y Universidad de Michigan (UofM) Registro de diabetes (n = 89)	El DKT es un instrumento de 23 ítems diseñado para evaluar el conocimiento del paciente sobre la diabetes en relación con la dieta, el ejercicio, los niveles de glucosa en sangre y las pruebas y las actividades de autocuidado. Los primeros 14 elementos se aplican a todos los pacientes y los 9 elementos restantes son relevantes solo para aquellos que usan insulina.	1-Validez de contenido 2-Consistencia interna 3- Prueba de hipótesis 4-Sensibilidad al cambio	1-Dudoso 2-Inadecuado 3- Dudoso 4- Dudoso	- α Cronbach para la prueba general = 0.77. - Se encontró una correlación significativa para la subescala de uso de insulina (r = 0.28, p = 0.03).
Mangla et al. (2018)	77 jóvenes indios con diabetes mellitus tipo 1 (entre 3 y 25 años de edad)	India: los pacientes fueron reclutados de la comunidad, habían recibido previamente educación en diabetes en hospitales.	El DKT de 34 ítems incluye preguntas sobre insulina y monitorización, terapia nutricional, ejercicio y su manejo, hipoglucemia y su manejo, hiperglucemia y manejo de días de enfermedad, viajes, ajuste de dosis y complicaciones. Puntuación DKT (puntuación máxima 100).	1-Validez de contenido 2-Consistencia interna	1-Inadecuado 2-Inadecuado	- α Cronbach = 0,87 para todo el cuestionario.

Estudio (autor y año)	Poplación / tipo de DM	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Michigan Diabetes Knowledge Test (DKT)						
Hasan et al. (2019)	10 diabéticos de habla árabe	Emiratos Árabes Unidos: Universidad de Ajman (AU). Los pacientes fueron reclutados en una de las clínicas locales para el cuidado de la diabetes.	La herramienta Simplified Diabetes Knowledge Test (SDKT) utilizada para esta adaptación consiste en una evaluación del conocimiento de la diabetes que contiene 20 ítems en formato Verdadero / Falso, 18 de los cuales pertenecen a la conciencia general sobre la diabetes, mientras que los otros dos se refieren al uso de insulina. Los puntajes de logro en el DKT simplificado se calculan como la proporción de respuestas correctas. Se considera una prueba rápida, ya que tarda unos 15 minutos en completarse. Es una versión simplificada de la escala original DKT.	1-Validez de contenido 2- Validez transcultural / invariancia de medición	1-Dudoso 2-Inadecuado	

Estudio (autor y año)	Poplación / tipo de DM	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Michigan Diabetes Knowledge Test (DKT)						
Idiz et al. (2020)	296 sujetos diabéticos tipo 1 y 2 que utilizan insulina	Turkia: Clínica ambulatoria de Pacientes ambulatorios	Adaptación turca de DKT-2. Consta de 23 preguntas de opción múltiple. Las primeras 14 preguntas miden el nivel general de conocimientos sobre diabetes. En las últimas 9 preguntas se evalúa el nivel de conocimiento sobre el uso de insulina. Si bien las primeras 14 preguntas se pueden aplicar a todos los adultos con diabetes tipo 1 y 2, las últimas 9 preguntas son relevantes solo para aquellos que usan insulina. A cada ítem se le asigna una puntuación de 1 para una respuesta correcta y 0 para una respuesta incorrecta. 23 elementos correctos dan un 100% de respuestas correctas. Tarda aproximadamente 15 minutos en completarse.	1-Validez de contenido 2-Consistencia interna 3- Validez transcultural / invariancia de medición 4- Error de medición 5-Fiabilidad 6-Prueba de hipótesis 7-Sensibilidad al cambio	1-Dudoso 2-Dudoso 3-Dudoso 4-Inadecuado 5-Inadecuado 6-Dudoso 7-Dudoso	- α Cronbach para prueba completa = 0.70. - Los valores de fiabilidad test-retest fueron 0.76 y 0.87 (p <0.001) para la primera y la segunda parte, respectivamente. - No hubo diferencia estadísticamente significativa entre los grupos en términos de tipo y duración de la diabetes, nivel de HbA1c y nivel de educación de la diabetes según las subescalas del nivel de conocimiento de la prueba general DKT2 y el uso de insulina de los participantes (p > 0.05)

Estudio (autor y año)	Poplación / tipo de DM	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Diabetes Knowledge Questionnaire (DKQ)						
García et al. (2001)	502 mexicanoamericanos adultos con diabetes tipo 2	México: la muestra se reclutó como parte del Estudio de educación sobre la diabetes del condado de Starr	El DKQ original es un instrumento de 60 ítems, las posibles opciones de respuesta son 1) Sí, 2) No y 3) No lo sé. Los elementos se puntúan como correctos o incorrectos, y los elementos correctos se suman para obtener una puntuación total. La administración de la prueba tarda 30 min.	1-Validez de contenido 2-Consistencia interna 3- Validez transcultural / invariancia de medición 4- Validez de criterio 5-Sensibilidad al cambio	1-Inadecuado 2-Inadecuado 3-Inadecuado 4- Inadecuado 5- Inadecuado	- α Cronbach = 0.83. - Las puntuaciones en las versiones de 60 y 24 ítems estaban bien correlacionadas ($r = 0.85$, $P < 0.001$).

Estudio (autor y año)	Poplación / tipo de DM	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Diabetes Knowledge Questionnaire (DKQ)						
García et al. (2001)	Como anteriormente	Como anteriormente	El Diabetes Knowledge Questionnaire-24 (DKQ-24) se compone de 24 preguntas sobre conocimientos básicos de la enfermedad (10 ítems), control glucémico (7 ítems) y prevención de complicaciones (7 ítems). Las preguntas son cerradas, con opciones de respuesta "sí, no y no sé", obteniendo 1 punto por respuesta correcta y la puntuación final es la suma total de puntos. Dado que no encontramos un punto de corte recomendado, establecimos como "conocimiento adecuado" a aquella persona con una puntuación superior a 13 o superior al 55% de respuestas correctas y "conocimiento inadecuado" a aquella con una puntuación inferior a 13.	Como anteriormente	Como anteriormente	- α Cronbach =0.78. - Las puntuaciones en las versiones de 60 y 24 ítems estaban bien correlacionadas ($r = 0.85$, $P < 0.001$).

Estudio (autor y año)	Poplación / tipo de DM	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Diabetes Knowledge Questionnaire (DKQ)						
Ahmad et al. (2010)	17 sujetos de Orang Asli que no son diabéticos	Malasia: población indígena	El Cuestionario Orang Asli (OA-DKQ) consta de 3 categorías; a) conocimientos generales sobre salud y enfermedades no transmisibles b) conocimientos específicos sobre diabetes (etiología, prevención, factores de riesgo) y c) comportamientos preventivos (peso y dieta, actividad física, tipos de tratamiento). Los elementos del OA-DKQ se limitaron a 12 y 89 subelementos para permitir una recopilación de datos concisa. Las opciones de respuesta para los diferentes ítems son "sí", "no" o "no sé". El cuestionario tarda entre 25 y 30 minutos en completarse.	1-Validez de contenido 2-Consistencia interna 3-Fiabilidad 4- Prueba de hipótesis 5-Sensibilidad al cambio	1-Dudoso 2-Inadecuado 3-Inadecuado 4- Inadecuado 5- Inadecuado	- α Cronbach = 0.806. - Según el ICC > 0.60 o el valor constante generado, el 62% (55 de 89) de los subelementos del grupo de control están altamente correlacionados (ICC > 0.60). En el grupo de casos, el 38% de los subelementos están altamente correlacionados (ICC > 0.60). Al comparar el ICC de las diferentes preguntas categóricas; relacionados con el conocimiento y relacionados con el comportamiento; para el grupo de control el 63% (35 de 56) tenía ICC > 0.60 y con respecto a los sub-elementos de conducta, el 60% (20 de 33) tenía ICC > 0.60. Por el contrario, en el grupo Orang Asli; con respecto a los sub-ítems de conocimiento, el 32% (18 de 56) tenían ICC > 0.60 y con respecto a los sub-ítems de comportamiento, el 48% (20 de 33) tenían ICC > 0.60.

Estudio (autor y año)	Poplación / tipo de DM	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Diabetes Knowledge Questionnaire (DKQ)						
Bukhsh et al. (2017)	130 pacientes con diabetes tipo 2	Pakistán: Akhuwat Hospital Lahore y Awan Medical Complex Lahore	El Diabetes Knowledge Questionnaire-24 (DKQ-24 se compone de 24 preguntas sobre conocimientos básicos de la enfermedad (10 ítems), control glucémico (7 ítems) y prevención de complicaciones (7 ítems). Las preguntas son cerradas, con opciones de respuesta "sí, no y no sé", obteniendo 1 punto por respuesta correcta y la puntuación final es la suma total de puntos. Como no encontramos un punto de corte recomendado, establecemos como "conocimiento adecuado" a aquella persona con una puntuación mayor a 13 o superior al 55% de respuestas correctas y "conocimiento inadecuado" a aquella con una puntuación menor a 13.	1-Consistencia interna 2- Validez transcultural / invariancia de medición 3- Prueba de hipótesis 4-Sensibilidad al cambio	1-Inadecuado 2-Inadecuado 3- Dudoso 4- Dudoso	- α Cronbach = 0.702. - Los resultados del análisis de varianza unidireccional (ANOVA) revelaron una diferencia estadísticamente significativa ($P < 0.05$) entre las puntuaciones de DKQ de los pacientes con "buen control glucémico" ($HbA1c \leq 7.5\%$), "control glucémico medio" ($HbA1c 7.6-8.9\%$) y "control glucémico deficiente" ($HbA1c \geq 9.0\%$) [$F(2,127) = 5.336, P = 0.006$].

Estudio (autor y año)	Poplación / tipo de DM	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Diabetes Knowledge Questionnaire (DKQ)						
Sami et al. (2017)	132 pacientes con diabetes mellitus tipo 2	Arabia Saudita	El cuestionario DKQ utilizado en esta investigación consta de 20 preguntas de opción múltiple. Las preguntas evalúan el conocimiento dietético de un paciente en cuanto a carbohidratos, lípidos, proteínas, tipo de alimentos y opciones de alimentos. Las respuestas están codificadas como 1 = correcto y 0 = incorrecto, y no lo sé. La puntuación varía de 0 a 20; un nivel de conocimiento dietético más alto se indica con una puntuación más alta. Se considera que un puntaje <50% presenta un conocimiento dietético pobre, un puntaje entre 50% -75% obtiene un buen conocimiento dietético y un puntaje > 75% se considera que tiene un conocimiento dietético adecuado.	1-Validez de contenido 2-Validez estructural 3-Consistencia interna 4-Prueba de hipótesis 5- Sensibilidad al cambio	1-Dudoso 2-Dudoso 3-Dudoso 4- Inadecuado 5- Inadecuado	- El EFA se realizó en el cuestionario de "actitud dietética", dando como resultado una solución de 5 factores. El factor 1 comprendía los ítems 2, 8, 12 y 13. El factor 2 comprendía los ítems 1, 3, 6 y 1. El factor 3 estaba compuesto por los ítems 9, 14 y 16. El factor 4 estaba compuesto por los ítems 4 y 15. El factor 5 estaba compuesto por los ítems 5, 7 y 11. Los 5 factores fueron etiquetados como "selección de alimentos", "impacto en la salud", "elecciones saludables", "restricción de alimentos" y "categorización de alimentos". - α Cronbach = 0.869. - La correlación total para DKQ varió de $r = 0.364 - 0.588$. Todos los valores estaban por encima del valor de corte > 0.30.

Estudio (autor y año)	Poplación / tipo de DM	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Diabetes Knowledge Assessment Test (DKAT)						
Weeks et al. (2014)	150 pacientes con o sin diabetes.	Estados Unidos: 75 pacientes hospitalizados y 75 ambulatorios.	El DKAT inicial constaba de 49 ítems, se redujo a 32 ítems según criterios psicométricos. Está desarrollado para ser administrado fácilmente a lectores y no lectores limitados, así como a personas con capacidad cognitiva limitada. Consta de 26 elementos con respuestas verdadero / falso o "No sé" y 6 elementos con respuestas de múltiples opciones. El tiempo medio de administración es de 16 minutos.	1-Validez de contenido 2-Validez estructural 3-Consistencia interna 4-Prueba de hipótesis 5-Sensibilidad al cambio	1-Dudoso 2-Dudoso 3-Adecuado 4- Dudoso 5- Dudoso	- El CFA identificó 2 factores: "Complicaciones" y "Manejo de síntomas de riesgos". Se probaron 2 modelos: el modelo de 2 factores teóricamente preferido (Modelo A) y uno asumiendo que la escala era unidimensional (Modelo B). El ajuste del modelo se evaluó mediante la prueba de significancia de chi-cuadrado, CFI, TLI y la raíz cuadrada media ponderada residual (WRMR). Los valores de WRMR inferiores a 1.0 indican un buen ajuste. Los valores de CFI y TLI superiores a 0.90 pueden indicar un ajuste adecuado. - α Cronbach = 0.82. - Los pacientes con diabetes obtuvieron puntuaciones DKAT significativamente más altas que los pacientes sin diabetes ($p = 0.01$).

Estudio (autor y año)	Poplación / tipo de DM	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Personal Diabetes Questionnaire (PDQ)						
Cheng et al. (2016)	346 adultos con diabetes tipo 2	China: Se reclutó a pacientes de cuatro hospitales terciarios afiliados a universidades en Xi'an.	El PDQ consta de 45 ítems que miden siete subescalas, que incluyen conocimiento sobre la dieta, toma de decisiones sobre la dieta, problemas de alimentación, barreras de la dieta, barreras de medicación, barreras de supervisión y barreras de ejercicio. Cada ítem fue calificado en una escala tipo Likert de 6 puntos con un rango posible de 1 (nunca) a 6 (una o más veces al día). Los puntajes altos en las subescalas de conocimiento de la dieta y toma de decisiones indican altos niveles de conocimiento de la dieta y habilidades para la toma de decisiones; las puntuaciones altas en las subescalas relacionadas con las barreras indican niveles altos de barreras en los dominios de autogestión correspondientes.	1-Validez de contenido 2-Validez estructural 3-Consistencia interna 4- Validez transcultural / invariancia de medición 5-Fiabilidad 6-Prueba de hipótesis 7-Sensibilidad al cambio	1-Dudoso 2- Dudoso 3-Inadecuado 4- Inadecuado 5- Inadecuado 6- Dudoso 7- Dudoso	- Se utilizó el EFA para explorar la validez de constructo factorial del C-PDQ obteniendo una solución de 2 factores etiquetados como; "Barreras situacionales" y "Barreras psicológicas". - El α Cronbach de las subescalas del C-PDQ osciló entre 0.61 y 0.83 para los pacientes que utilizan insulina y 0.69 a 0.89 para los pacientes que no la recibían, respectivamente. - Los ICC de 2 semanas para las subescalas oscilaron entre 0.73 y 0.96 para los pacientes que utilizan insulina y 0.75 a 0.96 para pacientes que no reciben insulina, respectivamente. - El criterio y la validez convergente fueron apoyados por correlaciones significativas de las subescalas del C-PDQ con los valores de HbA1c y las subescalas paralelas en el C-SDSCA, respectivamente.

Estudio (autor y año)	Poplación / tipo de DM	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Personal Diabetes Questionnaire (PDQ)						
Stetson et al. (2011)	790 adultos (205 con tipo 1 y 585 con diabetes tipo 2)	Estados Unidos: los participantes fueron reclutados mientras esperan sus citas clínicas programadas de rutina	El PDQ consta de 49 ítems, incluyen conocimientos y habilidades dietéticas, toma de decisiones, problemas alimentarios, barreras dietéticas, problemas en la medicación utilizada, barreras de medicación, supervisión y de ejercicio. Cada ítem se califica en una escala tipo Likert de 6 puntos con un rango posible de 1 (nunca) a 6 (una o más veces al día). Los puntajes altos en las subescalas de conocimiento de la dieta y toma de decisiones indican altos niveles de conocimiento de la dieta y habilidades para la toma de decisiones; las puntuaciones altas en las subescalas relacionadas con las barreras indican niveles altos de dificultad en los dominios de autogestión.	1-Consistencia interna 2-Prueba de hipótesis 3-Sensibilidad al cambio	1-Inadecuado 2-Dudoso 3-Dudoso	- α Cronbach osciló entre un máximo de 0.834 para barreras dietéticas para el grupo de DM2 que usaba insulina y un mínimo de 0.650 para barreras de ejercicio para el grupo de DM1. - Las subescalas demostraron asociaciones significativas con el IMC ($p \leq 0.001$) y la HbA1c ($p \leq 0.001$).

Estudio (autor y año)	Poplación / tipo de DM	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Diabetes Mellitus Knowledge (DMK)						
Sami et al. (2017)	132 pacientes con diabetes mellitus tipo 2	Arabia Saudita	El cuestionario DMK utilizado en esta investigación es la versión modificada / extendida del cuestionario DKQ desarrollado por Garcia et al. La versión original consta de 24 preguntas. La nueva versión del cuestionario DMK utilizada en esta investigación consta de 30 preguntas que se utilizan para evaluar con mayor detalle la "comprensión de la diabetes, los síntomas y el conocimiento, las complicaciones, la alimentación y el ejercicio, la medicación, la prevención y las investigaciones". Las respuestas se codifican como 1 = sí y 0 = no, y no lo sé.	1-Validez de contenido 2-Validez estructural 3-Consistencia interna 4-Prueba de hipótesis 5- Sensibilidad al cambio	1-Dudoso 2-Dudoso 3-Dudoso 4- Inadecuado 5- Inadecuado	- El EFA se realizó en el cuestionario de "actitud dietética", dando como resultado una solución de 5 factores con valores propios superiores a 1. El factor 1 comprendía los ítems 2, 8, 12 y 13. El factor 2 comprendía los ítems 1, 3, 6 y 1. El factor 3 estaba compuesto por los ítems 9, 14 y 16. El factor 4 estaba compuesto por los ítems 4 y 15. El factor 5 estaba compuesto por los ítems 5, 7 y 11. Los 5 factores fueron etiquetados como "selección de alimentos", "impacto en la salud", "elecciones saludables", "restricción de alimentos" y "categorización de alimentos". - α Cronbach = 0.891. - La correlación total para DMK varió de $r = 0.358 - 0.529$. Todos los valores estaban por encima del valor de corte > 0.30 .

Estudio (autor y año)	Poplación / tipo de DM	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Conocidiet-Diabetes						
Brito et al. (2019)	64 pacientes con diabetes tipo 1 y 2	España: Atención Primaria	Conocidiet-Diabetes se autoadministra, aunque también puede ser heteroadministrado. Consta de 88 ítems con respuestas verdadero/falso y tiene como objetivo medir el nivel de conocimiento dietético que tienen los pacientes con diabetes mellitus que acuden a las consultas de Atención Primaria, aunque también se podría poner a prueba en otros entornos. Está organizado por áreas de conocimiento en relación a la dieta, generado a partir de los 22 indicadores del NOC 1802, con 4 ítems para cada indicador.	1-Validez de contenido 2-Error de medición 3- Prueba de hipótesis 4-Sensibilidad al cambio	1-Adecuado 2-Inadecuado 3- Dudoso 4- Dudoso	- El grado de respuesta correcta fue de 25.0% y 100%. La correlación confirmada entre el grado de cumplimentación y el porcentaje de aciertos sugiere, en presencia de algunos ítems sin respuesta, que el motivo de su no cumplimentación fue la falta de conocimiento por parte de los pacientes.

Estudio (autor y año)	Poplación / tipo de DM	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Diabetes-related Nutrition Knowledge (DRNK)						
Han et al. (2019)	183 participantes: un grupo de estudiantes de nutrición (n = 21), un grupo de individuos con un estilo de vida libre con DM2 (n = 62) y otro grupo de estudiantes de nutrición (n = 100)	Singapur: los pacientes fueron reclutados de Diabetes Singapore, una organización de bienestar voluntaria para pacientes diabéticos y sus familias.	DRNK-Q se compone de 4 secciones y 27 preguntas. Todas las preguntas poseen opciones múltiples de respuesta (4 opciones de respuesta y una opción para "No sé"). Una respuesta correcta recibe 1 punto, mientras que una incorrecta o "No sé", recibe 0 puntos. Se puede completar el cuestionario en 10 minutos.	1-Validez de contenido 2-Consistencia interna 3-Prueba de hipótesis 4-Sensibilidad al cambio	1-Dudoso 2-Inadecuado 3-Dudoso 4-Dudoso	- α Cronbach = 0.66, 0.79, 0.56 y 0.78 para las secciones 1, 2, 3 y 4, respectivamente. - ICC = 0.82–0.84 - El grado de concordancia entre una prueba y una repetición de la prueba fue de 1 a -1. Un buen rango para ICC sería de 0.70 a 1. Los resultados demuestran un grado aceptable de acuerdo para cada uno de los grupos de prueba después de un intervalo de un mes.

Abreviaturas: (DM) Diabetes Mellitus; (DPKT) Diabetes Patient Knowledge Test; (DKN) Diabetes Knowledge Assessment Scale; (DKT) Diabetes Knowledge Test; (MDPH) Michigan Department of Public Health; (PAID) Problem area in diabetes scale; (HbA1c) Hemoglobina glicosilada; (DES) Diabetes Empowerment Scale; (MDKT) Michigan Diabetes Knowledge Test; (ICC) Intraclass Correlation Coefficient; (RDKS) Revised Diabetes Knowledge Scale; (MDKS) Michigan Diabetes Knowledge Scale; (DKT2) Diabetes Knowledge Test 2; (SDKS) Simplified Diabetes Knowledge Scale; (RVEEH) Royal Victorian Eye and Ear Hospital; (PSI) Person Separation Index; (PR) Person Reliability; (NDSS) National Diabetes Service Scheme; (UofM) University of Michigan's; (AU) Ajman University; (SDKT) Simplified Diabetes Knowledge Test; (DKQ) Diabetes Knowledge Questionnaire; (DKQ-24) Diabetes Knowledge Questionnaire- 24; (OA-DKQ) Orang Asli Diabetes Knowledge Questionnaire; (ANOVA) Analysis of Variance; (EFA) Análisis Factorial Exploratorio; (DKAT) Diabetes Knowledge Assessment Test; (CFA) Análisis Factorial Confirmatorio; (CFI) comparative fit index; (TLI) Tucker- Lewis fit index; (WRMR) Weighted Root Mean Square Residual; (PDQ) Personal Diabetes Questionnaire; (C-PDQ) Chinese - Personal Diabetes Questionnaire; (C-SDSCA) Chinese- Summary of Diabetes Self-Care Activities; (IMC) Índice de Masa Corporal; (DMK) Diabetes Mellitus Knowledge Questionnaire; (NOC) Nursing Outcomes Classification; (DRNK) Diabetes-related Nutrition Knowledge.

4.3 Marcos teóricos de los instrumentos

Los marcos teóricos de los instrumentos que evalúan el conocimiento dietético en DM tipo 1 o 2 no se informaron para todos los instrumentos. En este contexto, la mayoría de las herramientas se basaron en las recomendaciones de contenido de estándares para diabetes, revisiones de literatura y paneles de expertos (4,37–39) (tabla 7). Solo el Conocidiet-Diabetes se basó en la Teoría de Rango Medio para el Dominio del Automanejo de la Diabetes (40). Esta premisa une la teoría de la integración organísmica con la teoría del dominio con el fin de comprender cómo interactúan la motivación y el dominio para generar prácticas favorables en el contexto del cuidado de la DM. En este sentido, el conocimiento dietético en DM forma parte del proceso de adquisición del dominio del autocuidado.

El bajo conocimiento dietético puede percibirse como un factor estresante que evita el manejo de la enfermedad y reduce la motivación para mejorar los resultados del paciente.

Tabla 7. Marco teórico y proceso de desarrollo del instrumento.

Instrumento	Autor	Marco teórico y proceso de desarrollo del instrumento
Diabetes Patient Knowledge Test (DPKT)	Hess et al. (1983)	Desarrollo de instrumentos basado en expertos (médicos, enfermeras y dietistas) del Centro de Investigación y Capacitación en Diabetes de Michigan.
Diabetes Knowledge Assessment (DKN)	Dunn et al. (1984)	Marco teórico proporcionado por una búsqueda bibliográfica y un panel de expertos (médicos, enfermeras, dietistas y educadores en diabetes) de la Clínica de Diabetes del Royal Prince Alfred Hospital.
Michigan Diabetes Knowledge Test (DKT)	Fitzgerald et al. (1998)	Desarrollo de instrumentos basado en expertos (diabetólogos, dietistas, enfermeras, especialistas en educación y psicólogos) del Centro de Investigación y Capacitación en Diabetes de Michigan.
Diabetes Knowledge Questionnaire (DKQ)	García et al. (2001)	Marco teórico basado en las recomendaciones de contenido de los Estándares Nacionales para Programas de Educación de Pacientes con Diabetes, y panel de expertos (enfermeros experimentados e investigadores familiarizados con temas relacionados con la diabetes de los mexicoamericanos).
Diabetes Knowledge Assessment Test (DKAT)	Weeks et al. (2014)	El marco conceptual para el desarrollo de los ítems del instrumento estuvo representado por áreas de contenido seleccionadas para el conocimiento básico de la diabetes enseñado en Educación para el autocontrol de la diabetes como se describe en el estándar 6 de los Estándares nacionales para la educación para el autocontrol de la diabetes de la Asociación Estadounidense de Diabetes (ADA).
Personal Diabetes Questionnaire (PDQ)	Cheng et al. (2016)	El instrumento fue desarrollado por un equipo de investigación multidisciplinario basado en los estándares y puntos de referencia centrados en el paciente recomendados por la Asociación Estadounidense de Educadores en Diabetes y Diabetes Australia.
Diabetes Mellitus Knowledge (DMK)	Sami et al. (2017)	Este instrumento es la versión modificada del cuestionario DKQ que fue desarrollado por un equipo de investigación multidisciplinario basado en los estándares y puntos de referencia centrados en el paciente recomendados por la Asociación Estadounidense de Educadores en Diabetes y Diabetes Australia.
Conocidiet-Diabetes	Brito et al. (2019)	El instrumento se basó en el marco de la Teoría de rango medio para el dominio del autocontrol de la diabetes. El desarrollo del cuestionario se basó en el resultado NOC "Conocimiento: Dieta prescrita" (1802).
Diabetes-related Nutrition Knowledge (DRNK)	Han et al. (2019)	Marco teórico proporcionado por una búsqueda bibliográfica y un panel de expertos (tres dietistas con experiencia en asesoramiento dietético de pacientes con DM de Singapur, un académico con experiencia en investigación de diabetes y un psicólogo con experiencia en asesoramiento a pacientes con diabetes en el contexto de Singapur).

4.4 Propiedades psicométricas

Las principales propiedades evaluadas fueron la validez de contenido, la consistencia interna, la prueba de hipótesis y la sensibilidad al cambio. Los resultados de la lista de verificación COSMIN de cuatro puntos de instrumentos para medir el estado nutricional en pacientes con DM tipo 1 y/o 2 se muestran en la tabla 8. La tabla 9 muestra los datos sobre los instrumentos para medir el conocimiento de la diabetes utilizando el método GRADE.

4.4.1 *Diabetes Patient Knowledge Test (DPKT)*

Las propiedades psicométricas de este cuestionario fueron evaluadas en un único estudio (18). La validez de contenido, la validez estructural y la consistencia interna se calificaron como "inadecuadas". Para la validez estructural se identificaron 5 factores que fueron etiquetados como: "carbohidratos", "azúcar en sangre", "básicos", "intercambios de alimentos" y "administración de insulina". La puntuación de consistencia interna fue de 0.89. Para la prueba de hipótesis, las correlaciones factor-ítem de las dos formas de la prueba de conocimientos sobre diabetes para DPKT oscilaron entre $r = 0.20-0.67$ para el formulario A y $r = 0.14-0.63$ para el formulario B.

4.4.2 *Diabetes Knowledge Assessment Scale (DKN)*

DKN fue evaluado en solo uno de los estudios (19). La validez de contenido fue calificada como "adecuada", debido a que la mayoría de los ítems incluidos reflejaban adecuadamente el constructo que se pretendía medir. La puntuación de consistencia interna fue "inadecuada" con un valor alfa de Cronbach = 0.82. Para la prueba de hipótesis, las correlaciones obtenidas entre sí para las 3 versiones de dicho cuestionario (DKNA, DKNB y DKNC) fueron en todo caso superiores a $r = 0.90$.

4.4.3 *Michigan Diabetes Knowledge Test (DKT)*

DKT (1,20-29) estuvo presente en 11 de los 22 estudios y se evaluaron prácticamente todas las propiedades psicométricas, excepto el error de medición y la validez de criterio. La puntuación total para las 7 propiedades psicométricas evaluadas fue "inadecuada". Para la validez de contenido, la propiedad psicométrica más importante, la calidad metodológica fue calificada como "inadecuada" o "dudosa" para los 7 estudios en los que fue evaluada (21,22,25-29), debido principalmente a la falta de valoración de esta propiedad por parte de profesionales y de pacientes con DM.

La validez estructural obtuvo una calificación de "inadecuada" o "dudosa". Para el estudio de Sigurdardottir et al. (21), el análisis factorial reveló 2 factores: (a) angustia relacionada con vivir con DM y (b) angustia relacionada con el manejo de la DM. Para el estudio de Collins et al. (23), se realizó un análisis Rasch de SDKS adoptando una versión bidimensional. Finalmente, para Fenwick et al. (24), el análisis factorial obtuvo una escala unifactorial.

Para la consistencia interna se obtuvo un alfa de Cronbach entre 0.60 (1) y 0.94 (21), siendo calificado como "inadecuado" o "dudoso" para los 9 estudios en los que se evaluó (1,20-23,25-27,29). Se midió la fiabilidad test-retest para 3 estudios, resultando una puntuación de "inadecuada" con valores ICC = 0.894 (22), CCI = 0.90 (25), ICC = 0.76 e ICC = 0.87 para la primera parte del cuestionario (test general) y la segunda parte de este (uso de insulina), respectivamente (29). La prueba de hipótesis estuvo entre "inadecuada" y "dudosa" para la mayoría de los estudios (21-24,26,29), salvo uno de los artículos, en los que la calificación fue "adecuada", siendo el rango de correlación total entre su diferentes ítems de $r = 0.22$ a $r = 0.53$ (20).

4.4.4 *Diabetes Knowledge Questionnaire (DKQ)*

Este instrumento se incluyó en 4 estudios (30-33). Se evaluaron todas sus propiedades psicométricas, excepto el error de medición. La validez de contenido fue calificada como "inadecuada" o "dudosa" para 3 de los 4 estudios (30,31,33) debido principalmente a la valoración de esta propiedad por parte de profesionales y pacientes con DM.

La validez estructural solo se evaluó en un estudio (33), siendo “dudosa”. Para ello, se llevó a cabo un análisis factorial exploratorio (EFA), resultando una solución de 5 factores: “selección de alimentos”, “impacto en la salud”, “opciones saludables”, “restricción de alimentos” y “categorización de alimentos”. La consistencia interna se calificó como “inadecuada” o “dudosa”, obteniendo un valor alfa de Cronbach entre 0.702 (32) y 0.869 (33). En solo uno de los estudios (30) se comparó el DKQ-24 (versión reducida del original) con el DKQ-60 (versión original), por lo que se evaluó la propiedad psicométrica de validez de criterio, obteniendo la calificación de “inadecuada”. La correlación total para las versiones de 60 y 24 ítems fue de $r = 0.85$ (30).

En cuanto a los 3 estudios restantes, la prueba de hipótesis fue calificada como “inadecuada” en 2 de ellos (31,33) y “adecuada” en el otro (32). Para Ahmad et al. (31), los subítems del grupo control estaban altamente correlacionados entre sí para las personas con DM. En Bukhsh et al. (32), la correlación total de DKQ con respecto al total de sus 24 ítems varió de $r = -0.194$ a $r = 0.50$, mientras que la correlación DKQ versus HbA1c osciló de $r = -0.20$ a $r = 0.27$. Para el último de los estudios, Sami et al. (33), la correlación total de DKQ entre sus diferentes ítems fluctuó entre $r = 0.358-0.529$.

4.4.5 Diabetes Knowledge Assessment Test (DKAT)

DKAT (10) se incluyó en un solo estudio. Tanto la validez de contenido como la validez estructural fueron calificadas como “dudosas”. En el caso de la validez de contenido, por la valoración de dicha propiedad por parte de profesionales y pacientes con DM; en el caso de la validez estructural, consecuencia de la gestión de los valores perdidos. El análisis factorial confirmatorio (CFA) identificó 2 factores: “complicaciones” y “manejo de síntomas de riesgo”. La consistencia interna se calificó como “adecuada”, obteniendo un valor alfa de Cronbach de 0.82. En la prueba de hipótesis la correlación total en DKAT para los 32 ítems varió de $r = 0.26-0.6$.

4.4.6 Personal Diabetes Questionnaire (PDQ)

PDQ se evaluó en 2 de los estudios (34,35) para todas las propiedades psicométricas, excepto para el error de medición. Tanto la validez de contenido como la validez estructural se calificaron como "dudosas" en el estudio de Cheng et al (34). En el caso de la validez de contenido, por la valoración de dicha propiedad por parte de profesionales y pacientes con DM. En cuanto a la validez estructural, en este estudio se utilizó el EFA para examinar la validez del constructo factorial de PDQ, obteniéndose una solución de 2 factores, etiquetados como: "barreras situacionales" y "barreras psicológicas". Por otro lado, la falta de descripción de los porcentajes y el manejo de los valores perdidos contribuyeron a una menor calificación de calidad según COSMIN.

La consistencia interna fue "inadecuada" para ambos estudios, resultando un alfa de Cronbach de entre 0.61 y 0.83 para los pacientes que usaban insulina y de 0.69 a 0.89 para los pacientes que no la recibían (34); un alfa de Cronbach entre un máximo de 0.834 para las barreras de dieta en el grupo DM tipo 2 que usa insulina y un mínimo de 0.650 para las barreras de ejercicio en el grupo DM tipo 1 (35). La fiabilidad test-retest se midió en uno de los 2 estudios, siendo dicha puntuación "inadecuada", y el valor ICC = 0.73-0.96 para pacientes que utilizaban insulina e ICC = 0.75-0.96 para pacientes que no la recibían (34). Finalmente, la prueba de hipótesis mostró correlaciones significativas de las subescalas del PDQ con los valores de HbA1c y las subescalas paralelas del Chinese-Summary of Diabetes Self-Care Activities (C-SDSCA) (34). Por otro lado, se obtuvieron el mismo tipo de correlaciones para las subescalas de PDQ con el índice de masa corporal (IMC) ($p \leq 0.001$) y HbA1c ($p \leq 0.001$) (34), siendo "dudosa" la calidad metodológica resultante para dicha propiedad psicométrica.

4.4.7 Diabetes Mellitus Knowledge (DMK)

DMK fue evaluado en uno solo de los estudios (33). La validez de contenido fue catalogada como "dudosa", así como la validez estructural, donde se realizó un EFA, obteniéndose una solución de 5 factores. Estos 5 factores fueron etiquetados como: "selección de alimentos", "impacto en la salud", "opciones saludables", "restricción de alimentos" y "categorización de alimentos". Para este

estudio, el alfa de Cronbach fue de 0.891. En la prueba de hipótesis, la correlación total para DMK entre sus diferentes ítems osciló entre $r = 0.358-0.529$.

4.4.8 Conocidiet-Diabetes

El cuestionario fue evaluado para un único estudio (9). La validez de contenido fue calificada como "adecuada" porque la mayoría de los ítems incluidos reflejaban adecuadamente el constructo que se pretendía medir. Este fue el único estudio donde se evaluó el error de medición, obteniendo una calificación para la calidad metodológica de "inadecuada". Para la prueba de hipótesis, como dato más significativo, la correlación comprobada entre el grado de cumplimiento y el porcentaje de aciertos sugiere, ante la presencia de algunos elementos no contestados, que el motivo de su incumplimiento fue el desconocimiento por parte de los pacientes. Quizá esto se debió al grado de dificultad de estos elementos, más que a una falta de comprensión en la redacción de los enunciados.

4.4.9 Diabetes-related Nutrition Knowledge (DRNK)

DRNK fue evaluado únicamente en un estudio (36), siendo la validez de contenido "dudosa" debido a la valoración de dicha propiedad por parte de profesionales y pacientes con DM. La consistencia interna fue calificada como "inadecuada", obteniendo un alfa de Cronbach de 0.66, 0.79, 0.56 y 0.78 para los tramos 1, 2, 3 y 4, respectivamente. Para la prueba de hipótesis se obtuvo una calificación de "dudosa", con la correlación test-retest de $r = 0.82$ y $r = 0.84$ para el grupo de estudiantes de nutrición y el grupo de estudiantes de no nutrición, respectivamente.

Tabla 8. Calidad metodológica de los estudios y calidad de los resultados informados por propiedad de medición, instrumento y estudio.

Instrumento	Artículo	Validez de contenido		Validez estructural		Consistencia interna		Validez transcultural / invariancia de medición		Error de medición		Fiabilidad		Validez de criterio		Prueba de hipótesis		Sensibilidad al cambio	
		M	Q	M	Q	M	Q	M	Q	M	Q	M	Q	M	Q	M	Q	M	Q
Diabetes Patient Knowledge Test (DPKT)	Hess et al. (1983)	I	-	I	-	I	-									D	?	D	?
Diabetes Knowledge Assessment (DKN)	Dunn et al. (1984)	A	±			I	+									D	?	D	?
Diabetes Knowledge Test (DKT)	Fitzgerald et al. (1998)					I	-									A	+	D	+
	Sigurdardottir et al. (2008)	I	?	I	?	I	+	D	?							D	?	D	?
	Al-Qazaz et al. (2010)	D	?			I	+	D	?			I	+			D	?	D	?
	Collins et al. (2010) (RDKS)			D	-	I	-									I	?	I	?
	Collins et al. (2010) (SDKS)			D	-	I	-									I	?	I	?
	Fenwick et al. (2013)			D	+											D	?	D	?

Instrumento	Artículo	Validez de contenido		Validez estructural		Consistencia interna		Validez transcultural / invariancia de medición		Error de medición		Fiabilidad		Validez de criterio		Prueba de hipótesis		Sensibilidad al cambio Artículo	
		M	Q	M	Q	M	Q	M	Q	M	Q	M	Q	M	Q	M	Q	M	Q
Diabetes Knowledge Test (DKT)	Alhaiti et al. (2016)	D	?			I	+	D	?			I	-						
	Al Qahtani et al. (2016)					I	-	I	-										
	Fitzgerald et al. (2016)	D	?			I	+								D	+	D	+	
	Mangla et al. (2018)	I	+			I	+												
	Hasan et al. (2019)	D	±					I	?										
	Idiz et al. (2020)	D	±			D	-	D	?			I	+		D	?	D	?	
Diabetes Knowledge Questionnaire (DKQ)	García et al. (2001)	I	?			I	+	I	+					I	+			I	+
	Ahmad et al. (2010)	D	+			I	+					I	?		I	?	I	?	
	Bukhsh et al. (2017)					I	+	I	?						A	+	A	+	
	Sami et al. (2017)	D	-	D	?	D	+								I	+	I	+	

Instrumento	Artículo	Validez de contenido		Validez estructural		Consistencia interna		Validez transcultural / invariancia de medición		Error de medición		Fiabilidad		Validez de criterio		Prueba de hipótesis		Sensibilidad al cambio	
		M	Q	M	Q	M	Q	M	Q	M	Q	M	Q	M	Q	M	Q	M	Q
Diabetes Knowledge Assessment Test (DKAT)	Weeks et al. (2014)	D	+	D	?	A	+									D	+	D	+
Personal Diabetes Questionnaire (PDQ)	Cheng et al. (2016)	D	+ -	D	?	I	-	I	+			I	+			D	?	D	?
	Stetson et al. (2011)					I	-									D	?	D	?
Diabetes Mellitus Knowledge (DMK)	Sami et al. (2017)	D	-	D	?	D	+									I	+	I	+
Conocidiet-Diabetes	Brito et al. (2019)	A	+							I	?					D	?	D	?
Diabetes-related Nutrition Knowledge (DRNK)	Han et al. (2019)	D	+ -			I	-									D	?	D	?

M: Calidad metodológica del estudio calificado como I = Inadecuado, D = Dudoso, A = Adecuado, V = Muy bueno, casillas vacías = no informado.

Q: Calidad de los resultados calificados como + = calificación suficiente, ? = calificación indeterminada, ± = calificación inconsistente, - = calificación insuficiente, casillas vacías = no informado

Tabla 9. Resumen de hallazgos: calidad metodológica, calidad de los resultados y calidad de la evidencia reportada por propiedad e instrumento de medición.

Instrumento	Validez de contenido			Validez estructural			Consistencia interna			Validez transcultural / invariancia de medición			Error de medición			Fiabilidad			Validez de criterio			Prueba de hipótesis			Sensibilidad al cambio		
	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE
Diabetes Patient Knowledge Test (DPKT)	I	-	VL	I	-	VL	I	-	VL													D	?	L	D	?	L
Diabetes Knowledge Assessment (DKN)	A	±	M				I	+	VL													D	?	L	D	?	L
Michigan Diabetes Knowledge Test (DKT)	I	?	L	I	-	VL	I	-	VL	I	-	L				I	-	L				I	?	M	I	?	L
Diabetes Knowledge Questionnaire (DKQ)	I	-	VL	D	?	L	I	+	L	I	?	L				I	?	VL	I	+	VL	I	?	M	I	?	M
Diabetes Knowledge Assessment Test (DKAT)	D	+	L	D	?	L	A	+	M													D	+	L	D	+	L

Instrumento	Validez de contenido			Validez estructural			Consistencia interna			Validez transcultural / invariancia de medición			Error de medición			Fiabilidad			Validez de criterio			Prueba de hipótesis			Sensibilidad al cambio		
	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE
Personal Diabetes Questionnaire (PDQ)	D	±	L	D	?	L	I	-	L	I	+	VL				I	+	VL				D	?	M	D	?	M
Diabetes Mellitus Knowledge (DMK)	D	-	L	D	?	L	D	+	L													I	+	VL	I	+	VL
Conocidiet-Diabetes	A	+	M										I	?	VL							D	?	L	D	?	L
Diabetes-related Nutrition Knowledge (DRNK)	D	±	L				I	-	VL													D	?	L	D	?	L

M: Calidad metodológica del estudio calificado como I = Inadecuado, D = Dudoso, A = Adecuado, V = Muy bueno, casillas vacías = no informado

Q: Calidad de los resultados calificados como + = calificación suficiente, ? = calificación indeterminada, ± = calificación inconsistente, - = calificación insuficiente, casillas vacías = no informado.

QE: Calidad de la evidencia calificada como H = Alta, M = Moderada, L = Baja, VL = Muy baja.

5. DISCUSIÓN

Esta es la primera revisión sistemática en la que se identifican los instrumentos para medir el conocimiento dietético en pacientes con DM tipo 1 y/o 2, y en la que se evalúan las propiedades psicométricas y la calidad metodológica de los instrumentos, basándonos en la metodología COSMIN más actualizada (12,13). Se identificaron 9 instrumentos que fueron diseñados específicamente para evaluar el conocimiento dietético en pacientes diabéticos a través de la inclusión de 22 estudios.

Distintos factores como el ámbito y la población donde se utilizará el instrumento, sus dimensiones, el número de ítems y la evidencia mostrada en la evaluación de cada propiedad psicométrica, deben ser tenidos en cuenta por clínicos e investigadores para decidir qué instrumento es el más adecuado para medir este constructo. Además, se debe considerar el idioma y la cultura de la versión original del instrumento a emplear.

No hubo ningún instrumento para medir el conocimiento dietético, para el cual se evaluaran todas las propiedades psicométricas. Las propiedades evaluadas para la mayoría de los estudios fueron la validez de contenido, la consistencia interna, la prueba de hipótesis y la sensibilidad al cambio.

La calidad metodológica de los estudios para cada una de las propiedades psicométricas evaluadas no fue la esperada, resultando "inadecuada" o "dudosa" en prácticamente la totalidad de los instrumentos. En el caso de la validez de contenido, esto se debió a la falta de evaluación de la relevancia de los ítems en la población de estudio. En la validez estructural, a la falta de descripción de los porcentajes y manejo de los valores perdidos. En la consistencia interna, al CFA no realizado, a la consistencia interna no calculada para cada dimensión por separado o a la falta de descripción de los porcentajes y manejo de los valores perdidos. En la validez transcultural, a la ausencia de comparación entre los diferentes grupos tras emplear este análisis transcultural. En la fiabilidad y la validez de criterio, a la falta de descripción de los porcentajes y manejo de los valores perdidos. Y, finalmente, en la prueba de hipótesis, esto fue consecuencia de que las hipótesis no eran claras o no fueron formuladas *a priori*, pero era posible deducir lo que se esperaba, y a la falta de descripción del manejo de los valores perdidos.

DPKT (18), DKN (19), DKAT (10), DMK (33), Conocidiet-Diabetes (9) y DRNK (36) fueron evaluados en un solo estudio. Por lo tanto, se requieren estudios adicionales para evaluar las propiedades psicométricas de estos instrumentos.

Únicamente hubo 3 instrumentos en los que la calidad metodológica fue calificada como "adecuada": para la validez de contenido en el instrumento DKN (19) y Conocidiet-Diabetes (9), consecuencia de que la mayoría de los ítems incluidos en ambos estudios fueron un reflejo adecuado del constructo que se pretendía medir, y para la consistencia interna del cuestionario DKAT (10), resultado del tamaño de la muestra.

La calidad de la evidencia fue "baja" o "muy baja" para la mayoría de las propiedades de medida analizadas en todos los instrumentos incluidos en esta revisión sistemática. Teniendo en cuenta que la validez de contenido, la validez estructural y la consistencia interna se consideran las propiedades de medición más importantes (11,12), la validez de contenido estuvo presente en los 9 instrumentos incluidos en la revisión: DPKT (18), DKN (19), DKT (21,22,25–29), DKQ (30,31,33), DKAT (10), PDQ (34), DMK (33), Conocidiet-Diabetes (9) y DRNK (36). Para 2 de los instrumentos, DKN (19) y Conocidiet-Diabetes (9), la calidad de la evidencia fue calificada como "moderada", debido a la probabilidad de que la propiedad de medición real esté cerca de la estimación de la propiedad de medición, aun existiendo la posibilidad de que sea sustancialmente diferente. Por otro lado, para DKAT (10) y Conocidiet-Diabetes (9) la calidad de los resultados fue catalogada como "+", indicando que los ítems de la medida fueron relevantes, exhaustivos y comprensibles con respecto al constructo a medir.

Para la validez estructural, la calidad de la evidencia también fue "baja" o "muy baja" para 6 de los instrumentos: DPKT (18), DKT (21,23,24), DKQ (33), DKAT (10), PDQ (34) y DMK (33). Para la mitad de los instrumentos: DPKT (18), DKQ y DMK (33) se obtuvo una estructura de 5 factores, mientras que para la otra mitad: DKT (21), DKAT (10) y PDQ (34), fueron 2 factores resultantes.

Finalmente, la consistencia interna también obtuvo una calificación de "baja" o "muy baja" para prácticamente todos los instrumentos en los que se evaluó esta propiedad psicométrica: DPKT (18), DKN (19), DKT (1,20–23,25–27,29), DKQ (30–33), PDQ (34,35), DMK (33) y DRNK (36). La excepción fue la herramienta de conocimiento dietético DKAT (10), en la que la calificación para la

calidad de la evidencia fue “moderada”. Se obtuvo una calidad de los resultados calificada como “+” (alfa de Cronbach ≥ 0.70) para los instrumentos DKN (19), DKQ (30–33), DKAT (10) y DMK (33), mientras que para los instrumentos restantes (DPKT (18), DKT (1, 20, 23, 29), PDQ (34, 35) y DRNK (36)) fueron “-” (alfa de Cronbach < 0.70).

DKT (1,20–29) ha sido el cuestionario incluido en el mayor volumen de estudios (11 estudios en total) y el segundo junto con PDQ (34,35), en el que más características psicométricas se han evaluado (7 en total). El cuestionario con más características psicométricas evaluadas ha sido DKQ (8 en total) (30-33), que es también el cuestionario con mejor puntuación en cuanto a la calidad de la evidencia, junto con PDQ (34,35), seguido finalmente por DKT (1,20–29). Esto hace que estos 3 cuestionarios sean los más ampliamente validados y los más recomendados para evaluar el conocimiento dietético en personas con DM.

5.1 Fortalezas y limitaciones

Es la revisión sistemática más actual que identifica, evalúa y resume la evidencia con respecto a los instrumentos para medir el conocimiento dietético en pacientes con DM tipo 1 y 2. La revisión se llevó a cabo de acuerdo con los estándares COSMIN y PRISMA más actualizados. Por lo tanto, se utilizó un método apropiado, siguiendo las recomendaciones de los expertos (11,12,14). Si bien se realizó una búsqueda minuciosa en diversas bases de datos, es posible que en esta revisión no se hayan identificado todos los instrumentos disponibles para la medición de este constructo, dado que la estrategia de búsqueda se limitó únicamente a artículos publicados en inglés y español, por lo que es probable que se haya prescindido de trabajos de investigación sobre instrumentos desarrollados y empleados en otros idiomas.

6. CONCLUSIONES

Se han desarrollado diferentes instrumentos para la evaluación del conocimiento dietético en personas con DM. Dicha evaluación nos permitirá mejorar la atención en esta población, y los instrumentos de evaluación resultantes, servirán para facilitar el cribado, la evolución, orientar sobre las

decisiones de tratamiento y establecer intervenciones enfocadas a mejorar el manejo dietético en este tipo de pacientes con esta patología concreta.

Según la evidencia actual de todas las propiedades de medición de los instrumentos que evalúan el nivel de conocimiento dietético en DM tipo 1 o 2, DKT, DKQ y PDQ parecen ser los instrumentos recomendados. Aun así, se necesitan más estudios de alta calidad metodológica que evalúen las propiedades de medición de los instrumentos para valorar el conocimiento dietético en pacientes con DM.

REFERENCIAS BIBLIOGRÁFICAS

1. Al Qahtani L, Alqarni A, Mohamud MS, Masuadi E, Aldhubayee M. Michigan Diabetes Knowledge Test: translation and validation study of the Arabic version.
2. Saeedi P, Petersohn I, Salpea P, Malanda B, Karuranga S, Unwin N, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition. *Diabetes Res Clin Pract.* noviembre de 2019;157:107843.
3. Sinclair A, Saeedi P, Kaundal A, Karuranga S, Malanda B, Williams R. Diabetes and global ageing among 65–99-year-old adults: Findings from the International Diabetes Federation Diabetes Atlas, 9th edition. *Diabetes Research and Clinical Practice* [Internet]. 1 de abril de 2020 [citado 26 de octubre de 2020];162. Disponible en: [https://www.diabetesresearchclinicalpractice.com/article/S0168-8227\(20\)30137-6/abstract](https://www.diabetesresearchclinicalpractice.com/article/S0168-8227(20)30137-6/abstract)
4. Funnell MM, Brown TL, Childs BP, Haas LB, Hoseney GM, Jensen B, et al. National Standards for Diabetes Self-Management Education. *Diabetes Care* [Internet]. 1 de enero de 2008 [citado 14 de mayo de 2020];31(Supplement 1):S97-104. Disponible en: <http://care.diabetesjournals.org/cgi/doi/10.2337/dc08-S097>
5. Jalilian H, Pezeshki MZ, Janati A, Najafipour F, Imani A, Zozani MA, et al. Readiness for diet change and its association with diet knowledge and skills, diet decision making and diet barriers in type 2 diabetic patients. *Diabetes Metab Syndr.* octubre de 2019;13(5):2933-8.
6. Sami W, Alabdulwahhab KM, Ab Hamid MR, Alasbali TA, Alwadani FA, Ahmad MS. Dietary Knowledge among Adults with Type 2 Diabetes— Kingdom of Saudi Arabia. *International Journal of Environmental Research*

- and Public Health [Internet]. enero de 2020 [citado 6 de junio de 2020]; 17(3):858. Disponible en: <https://www.mdpi.com/1660-4601/17/3/858>
7. Davis E. Role of the diabetes nurse educator in improving patient education. *The Diabetes Educator*. 1990;16(1):36-8.
 8. Nikitara M, Constantinou CS, Andreou E, Diomidous M. The role of nurses and the facilitators and barriers in diabetes care: A mixed methods systematic literature review. *Behavioral Sciences*. 2019;9(6):61.
 9. Brito-Brito PR, Martín-García Á, Oter-Quintana C, Paloma-Castro O, Romero-Sánchez JM. Development and Content Validation of a NOC-Based Instrument for Measuring Dietary Knowledge in Patients with Diabetes: CoNOCidiet-Diabetes. *International Journal of Nursing Knowledge* [Internet]. 2020 [citado 14 de mayo de 2020];31(1):59-73. Disponible en: <https://onlinelibrary.wiley.com/doi/abs/10.1111/2047-3095.12243>
 10. Weeks DL, French BF, Davis AD, Towle LA. Development and validation of the Diabetes Knowledge Assessment Test for use in medical rehabilitation. *Disability and Rehabilitation* [Internet]. 24 de abril de 2015 [citado 14 de mayo de 2020];37(9):802-11. Disponible en: <https://doi.org/10.3109/09638288.2014.940427>
 11. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res* [Internet]. 1 de mayo de 2018 [citado 15 de mayo de 2020];27(5):1147-57. Disponible en: <https://doi.org/10.1007/s11136-018-1798-3>
 12. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res*. 1 de mayo de 2010;19(4):539-49.
 13. Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, de Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual*

- Life Res [Internet]. 1 de mayo de 2012 [citado 14 de mayo de 2020];21(4):651-7. Disponible en: <https://doi.org/10.1007/s11136-011-9960-1>
14. Moher D, Liberati A, Tetzlaff J, Altman DG, Altman D, Antes G, et al. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement (Chinese edition). *Journal of Integrative Medicine* [Internet]. 1 de septiembre de 2009 [citado 14 de mayo de 2020];7(9):889-96. Disponible en: <https://jhu.pure.elsevier.com/en/publications/preferred-reporting-items-for-systematic-reviews-and-meta-analyse-10>
 15. Terwee CB, Jansma EP, Riphagen II, de Vet HCW. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res.* octubre de 2009;18(8):1115-23.
 16. Samuel-Hodge CD, DeVellis RF, Ammerman A, Keyserling TC, Elasy TA. Reliability and validity of a measure of perceived diabetes and dietary competence in African American women with type 2 diabetes. *The Diabetes Educator.* 2002;28(6):979-88.
 17. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* [Internet]. 24 de abril de 2008 [citado 15 de mayo de 2020];336(7650):924-6. Disponible en: <https://www.bmj.com/content/336/7650/924>
 18. Hess GE, Davis WK. The validation of a diabetes patient knowledge test. *Diabetes Care.* 1983;6(6):591-6.
 19. Dunn SM, Bryson JM, Hoskins PL, Alford JB, Handelsman DJ, Turtle JR. Development of the diabetes knowledge (DKN) scales: forms DKNA, DKNB, and DKNC. *Diabetes care.* 1984;7(1):36-41.
 20. Fitzgerald JT, Funnell MM, Hess GE, Barr PA, Anderson RM, Hiss RG, et al. The reliability and validity of a brief diabetes knowledge test. *Diabetes care.* 1998;21(5):706-10.

21. Sigurdardottir AK, Benediktsson R. Reliability and validity of the Icelandic version of the problem area in diabetes (PAID) scale. *International journal of nursing studies*. 2008;45(4):526-33.
22. Al-Qazaz HK, Hassali MA, Shafie AA, Sulaiman SAS, Sundram S. The 14-item Michigan Diabetes Knowledge Test: translation and validation study of the Malaysian version. *Practical Diabetes International*. 2010;27(6):238-241a.
23. Collins GS, Mughal S, Barnett AH, Fitzgerald J, Lloyd CE. Modification and validation of the revised diabetes knowledge scale. *Diabetic Medicine*. 2011;28(3):306-10.
24. Fenwick EK, Xie J, Rees G, Finger RP, Lamoureux EL. Factors associated with knowledge of diabetes in patients with type 2 diabetes using the diabetes knowledge test validated with Rasch analysis. *PloS one*. 2013;8(12).
25. Alhaiti AH, Alotaibi AR, Jones LK, DaCosta C, Lenon GB. Psychometric evaluation of the revised Michigan diabetes knowledge test (v. 2016) in arabic: translation and Validation. *Journal of diabetes research*. 2016;2016.
26. Fitzgerald JT, Funnell MM, Anderson RM, Nwankwo R, Stansfield RB, Piatt GA. Validation of the revised brief diabetes knowledge test (DKT2). *The Diabetes Educator*. 2016;42(2):178-87.
27. Mangla P, Chopra A, Sudhanshu S, Bhatia E, Dabadghao P, Gupta S, et al. Validation of a diabetes knowledge test for Indian children, adolescents and young adults with type 1 diabetes mellitus. *Primary care diabetes*. 2019;13(3):283-7.
28. Hasan S, Mahameed S, AlHariri Y. Translation and culture adaptation of the Simplified Diabetes Knowledge Test, the Literacy Assessment for Diabetes and the Diabetes Numeracy Test. *Research in Social and Administrative Pharmacy*. 2019;
29. İDİZ C, ÇELİK S, BAĞDEMİR E, DIŞSİZ M, SATMAN İ. Turkish Adaptation of Michigan Diabetes Research and Training Center's Revised Diabetes

- Knowledge Test and Determination of Factors Affecting the Knowledge Level of Diabetic Individuals. *Turkish Journal of Endocrinology & Metabolism*. 2020;24(1).
30. Garcia AA, Villagomez ET, Brown SA, Kouzekanani K, Hanis CL. The Starr County Diabetes Education Study: development of the Spanish-language diabetes knowledge questionnaire. *Diabetes care*. 2001;24(1):16-21.
 31. Ahmad B, Ramadas A, Quek KF. The development and validation of diabetes knowledge questionnaire for the Indigenous population in Malaysia. *Med J Malaysia*. 2010;65(4):274-7.
 32. Bukhsh A, Lee SWH, Pusparajah P, Khan AH, Khan TM. Psychometric properties of the Urdu version of diabetes knowledge questionnaire. *Frontiers in public health*. 2017;5:139.
 33. Sami W, Ansari T, Butt NS, Ab Hamid MR. Psychometric evaluation of dietary habits questionnaire for type 2 diabetes mellitus. En: *Journal of Physics: Conference Series*. IOP Publishing; 2017. p. 012151.
 34. Cheng L, Leung DY, Wu Y-N, Sit JW, Yang M-Y, Li X-M. Psychometric properties of the modified personal diabetes questionnaire among Chinese patients with type 2 diabetes. *Evaluation & the health professions*. 2018;41(1):3-24.
 35. Stetson B, Schlundt D, Rothschild C, Floyd JE, Rogers W, Mokshagundam SP. Development and validation of The Personal Diabetes Questionnaire (PDQ): a measure of diabetes self-care behaviors, perceptions and barriers. *Diabetes research and clinical practice*. 2011;91(3):321-32.
 36. Han CY, Zheng X, Lee LF, Chan CG, Lee YQ, Zailani NA, et al. Development of a diabetes-related nutrition knowledge questionnaire for individuals with type 2 diabetes mellitus in Singapore. *Nutrition & Dietetics*. 2019;76(5):567-73.

37. Association AD. National Standards for Diabetes Patient Education Programs (National Diabetes Advisory Board, November 1983). *Diabetes Care* [Internet]. 1 de enero de 1984 [citado 16 de noviembre de 2020];7(1):XXXI-XXXV. Disponible en: <https://care.diabetesjournals.org/content/7/1/XXXI>
38. Etzwiler DD, Robb JR. Evaluation of programmed education among juvenile diabetics and their families. *Diabetes*. 1972;21(9):967-71.
39. Etzwiler DD. What the juvenile diabetic knows about his disease. *Pediatrics*. 1962;29(1):135-41.
40. Fearon-Lynch JA, Stover CM. A middle-range theory for diabetes self-management mastery. *Advances in Nursing Science*. 2015;38(4):330-46.

Artículo 3

Cancer-related fatigue measures in palliative care: a psychometric systematic review

ARTÍCULO ACTUALMENTE EN REVISIÓN EN EUROPEAN JOURNAL OF CANCER CARE.

CANCER-RELATED FATIGUE MEASURES IN PALLIATIVE CARE: A PSYCHOMETRIC SYSTEMATIC REVIEW

RESUMEN

Antecedentes: En los cuidados paliativos (CP), la prevención y el alivio de la fatiga se consideran objetivos cruciales en pacientes con cáncer. Por lo tanto, una medición adecuada de la fatiga relacionada con el cáncer (FRC) es esencial en CP. En este sentido, se necesitan instrumentos de medición sólidos que nos permitan evaluar este constructo adecuadamente en el entorno clínico y recopilar datos óptimos en investigación. Se han diseñado y validado varios instrumentos para evaluar este constructo. Sin embargo, aún se desconoce cuál de estos es el más adecuado para medir la FRC en CP.

Objetivos: Identificar todos los instrumentos utilizados para la evaluación de la FRC en CP y evaluar críticamente sus propiedades de medición.

Métodos: Se llevó a cabo una revisión sistemática de acuerdo con la metodología más actualizada de los Estándares basados en el COnsenso para la selección de INstrumentos de Medición del estado de salud (COSMIN). La estrategia de búsqueda se realizó siguiendo la metodología de Revisión por Pares de las Estrategias de Búsqueda Electrónica (PRESS). Las búsquedas se llevaron a cabo en Medline (a través de PubMed), Web of Science, Open Gray, Scielo, Cochrane, CINAHL y EMBASE. Los artículos fueron evaluados por 2 revisores independientes, de acuerdo con la lista de verificación de riesgo de sesgo de COSMIN y los datos de las propiedades de medición de cada estudio que cumplía con los criterios de inclusión fueron calificados de forma independiente por dos investigadores de acuerdo con la metodología COSMIN. Las calificaciones de calidad y la síntesis de la evidencia se realizaron utilizando el método de Calificación de Recomendaciones, Valoración, Desarrollo y Evaluación (GRADE) modificado.

Resultados: Se identificaron 5598 artículos en las diferentes bases de datos. En total, se incluyeron 57 estudios que describen 19 instrumentos. Las principales propiedades evaluadas fueron: consistencia interna, validez transcultural, prueba de hipótesis y sensibilidad al cambio. Edmonton Symptom Assessment System (ESAS), Problems and Needs in Palliative Care Questionnaire (PNPC) y European Organization for Research and Treatment of Cancer Quality of Life 15 items Questionnaire for Palliative Care (EORTC QLQ-C15-PAL) fueron los cuestionarios con mayor presencia en cuanto a número de estudios, de propiedades psicométricas evaluadas y más importantes, lo que hace que estos 3 cuestionarios sean los más ampliamente validados y los más recomendados para su uso en la evaluación de la FRC en CP. Para esta evaluación también se debe incluir Palliative Care Quality of Life Instrument (PQLI), ya que a pesar de haber sido usado en un solo estudio, presentó las mejores puntuaciones en cuanto a calidad metodológica y calidad de resultados para las propiedades psicométricas más significativas.

Conclusiones: Según la calidad metodológica de los resultados, ESAS, PNPC, EORTC QLQ-C15-PAL y PQLI son los instrumentos recomendados para evaluar FRC en CP.

Número de registro PROSPERO: CRD42020206783.

Palabras clave: fatiga, instrumentos, enfermería, cuidados paliativos, propiedades psicométricas, revisión sistemática.

ABSTRACT

Background: In palliative care (PC), the prevention and relief of fatigue are regarded as crucial goals in patients with cancer; therefore, the measurement of optimal cancer-related fatigue (CRF) in PC is essential. In this regard, sound measurement instruments are needed to allow us to evaluate this construct properly in the clinical setting and for gathering good research data. Several instruments have been designed and validated to assess this construct. However, it remains unknown which of these instruments is the most fit-for-purpose to measure CRF in PC.

Objectives: To identify all the instruments used for assessing CRF in PC and to critically appraise their measurement properties.

Methods: A systematic review was carried out according to the most up-to-date CONsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) methodology. The search strategy was performed following the Peer Review of Electronic Search Strategies (PRESS) guidelines. Searches were conducted in Medline (through PubMed), Web of Science, Open Gray, Scielo, Cochrane, CINAHL and EMBASE. Papers were evaluated by two independent reviewers according to the COSMIN risk of bias checklist, and the measurement properties data of each study that met the inclusion criteria were scored independently by 2 researchers according to the COSMIN methodology. Quality ratings and evidence synthesis were performed using the modified Grading of Recommendations, Assessment, Development and Evaluation (GRADE) approach.

Results: 5598 articles were identified in the different databases. In total, 57 studies describing 19 instruments were included in this study. The main properties evaluated were internal consistency, cross-cultural validity, hypotheses testing, and responsiveness. Edmonton Symptom Assessment System (ESAS), Problems and Needs in Palliative Care Questionnaire (PNPC) and European Organization for Research and Treatment of Cancer Quality of Life 15 items Questionnaire for Palliative Care (EORTC QLQ-C15-PAL) were the questionnaires with the greatest presence in terms of number of studies, with more evaluated and more important

psychometric properties resulting in these 3 questionnaires being the most widely validated and the most recommended for use in the evaluation CRF in PC. For this evaluation, Palliative Care Quality of Life Instrument (PQLI) should also be included, since despite having been used in a single study, it presented the best scores in terms of methodological quality and quality of results for the most important psychometric properties.

Conclusions: According to the quality methodological results, ESAS, PNPC, EORTC QLQ-C15-PAL and PQLI are the recommended instruments used for assessing CRF in PC.

PROSPERO registration number: CRD42020206783.

Keywords: fatigue, instruments, nursing, palliative care, psychometric properties, systematic review.

1. INTRODUCCIÓN

El cáncer es un problema cuya incidencia y prevalencia ha aumentado en los últimos años, siendo la segunda causa de muerte a nivel mundial (1). Los servicios de cuidados paliativos (CP) pueden tener un efecto positivo en los pacientes con cáncer avanzado y sus cuidadores (2,3), siendo relevantes para mejorar tanto los resultados del paciente y del cuidador como para el manejo de los síntomas producidos por la enfermedad (4).

Los pacientes con cáncer avanzado experimentan múltiples síntomas como dolor, disnea y fatiga, que afectan negativamente a su calidad de vida relacionada con la salud (CVRS) (5,6). En este contexto, la fatiga relacionada con el cáncer (FRC) es un síntoma frecuente y se considera como una de las consecuencias de la enfermedad más importantes para los pacientes (7). Sin embargo, este síntoma es ampliamente ignorado por los médicos cuando evalúan la gravedad y el tratamiento de la enfermedad (7).

La fatiga se puede describir como una abrumadora sensación de cansancio, falta de energía y sensación de agotamiento, y se reconoce como un síntoma complejo, con aspectos tanto físicos como psicológicos (8,9). Así, la experiencia de la fatiga es multidimensional: se manifiesta no solo como una sensación física, sino también de manera emocional y cognitiva (8,9). En este sentido, la evaluación de la fatiga es, por definición, subjetiva y está influenciada por factores físicos y psicológicos que interactúan entre sí (8). En general, la fatiga es un constructo difícil de medir y uno de los síntomas más temidos por los pacientes con cáncer avanzado (7,8). Por ello, los enfermeros juegan un papel crucial en la evaluación y alivio de la FRC (10).

Por un lado, la FRC es más severa, más persistente y más debilitante que "otros tipos de fatiga", como los causados por la falta de sueño o el sobreesfuerzo, y esta no se alivia con el sueño o el descanso adecuado. La FRC es multidimensional y puede tener manifestaciones físicas y emocionales que incluyen debilidad generalizada, disminución de la concentración, disminución de la motivación o el interés por participar en las actividades habituales y labilidad emocional (9).

Por otro lado, la fatiga es el síntoma más común e ignorado en pacientes de CP que tienen cáncer avanzado u otras enfermedades graves y/o potencialmente

mortales. La fatiga tiene importantes consecuencias físicas, psicosociales y económicas, tanto para los pacientes como para los cuidadores. Sin embargo, debido a su naturaleza subjetiva y causas multidimensionales, la evaluación y el tratamiento de esta en los entornos de CP puede ser compleja.

Las propiedades y los factores asociados a la fatiga difieren entre la población de pacientes (supervivientes, pacientes en tratamiento o pacientes con cáncer avanzado). Su intensidad y sus propiedades también pueden variar según la localización del cáncer y la modalidad de tratamiento, y es posible que los instrumentos validados en un solo tipo de cáncer o durante su tratamiento no se puedan generalizar a todos los pacientes con cáncer. Hay pocos instrumentos que tengan en cuenta a poblaciones con cáncer en particular; la mayoría están validados para una combinación de personas supervivientes, pacientes que reciben quimioterapia o radiación, y en aquellos con cáncer avanzado que ya no reciben terapia antitumoral (11).

La Fatigue Coalition ha estado discutiendo la importancia de una evaluación rutinaria de la FRC a través de medidas válidas (12). En este sentido, necesitamos herramientas que nos permitan medir este constructo para recopilar más evidencia, ya que la medición de la FRC se usa cada vez más en la investigación y en el ámbito clínico (13–15). En este contexto, elegir la herramienta más apropiada para evaluar la FRC es un desafío, y la elección debe depender de las propiedades de medición y la viabilidad de los instrumentos (16).

Aunque se han realizado varias revisiones sistemáticas para identificar los instrumentos que evalúan la FRC, no se ha publicado ninguna revisión metodológicamente sólida que contraste las propiedades de medición de todos los instrumentos que valoran la FRC en CP de acuerdo con los Estándares más actualizados basados en el COnsenso para la selección de la metodología de INstrumentos de Medición del estado de salud (COSMIN) (13–17).

Por lo tanto, el objetivo de esta revisión sistemática es doble: identificar todos los instrumentos que se han utilizado para evaluar la FRC en CP y evaluar críticamente las propiedades de medición de todos estos instrumentos.

2. MÉTODOS

2.1 Diseño

Se efectuó una revisión sistemática de las propiedades de medida de los instrumentos que evalúan la FRC en CP de acuerdo con las Directrices de Elementos de Informe Preferidos para Revisiones Sistemáticas y Protocolos de Metanálisis (PRISMA-P). El protocolo de esta revisión se inscribió en el Registro Prospectivo Internacional de Revisiones Sistemáticas (PROSPERO) (nº de registro CRD42020206783) (18).

La revisión sistemática se llevó a cabo de acuerdo con la guía COSMIN más actualizada (17).

2.2 Estrategia de búsqueda

Las búsquedas se llevaron a cabo en Medline (a través de PubMed), Web of Science, Open Gray, Scielo, Cochrane, CINAHL y EMBASE siguiendo la metodología de Revisión por Pares de las Estrategias de Búsqueda Electrónica (PRESS) (19) durante marzo de 2021 por 2 revisores independientes. La estrategia de búsqueda incluyó una combinación de los siguientes 5 aspectos según Terwee et al: 1) búsqueda del constructo AND 2) búsqueda de población AND 3) búsqueda de instrumentos AND 4) filtro para propiedades psicométricas AND 5) filtro de exclusión "NOT" (20).

Se utilizaron los siguientes términos clave para ejecutar esta búsqueda: "fatigue", "tiredness", "weakness", "cancer-related fatigue", "palliative care", "terminal care", "end-of-life care", "terminally ill", "advanced illness", "terminal illness", "end of life", "instrument", "measure", "questionnaire", "scale", "tool", "survey" and "test", adaptando esta estrategia a cada uno de los sistemas de búsqueda de las diferentes bases de datos.

Para encontrar el máximo número de herramientas posibles y sus diferentes versiones, no hubo límites en las fechas de publicación. La búsqueda se limitó a artículos publicados en inglés y español. Además, se revisaron las listas de referencias de los artículos incluidos para identificar otros documentos relevantes (tabla 10). El filtro de exclusión se limitó principalmente a tipos de publicaciones

y grupos de temas. La estrategia de búsqueda se llevó a cabo siguiendo la metodología COSMIN para revisiones sistemáticas de medidas de resultado informadas por el paciente (PROM).

Tabla 10. Estrategia de búsqueda.

Estrategia de búsqueda para Pubmed

- 1 "Fatigue"[MeSH] OR "Tiredness"[MeSH] OR "Weakness"[tiab] OR "Cancer related fatigue"[tiab]
- 2 "Palliative care"[MeSH] OR "terminal care"[MeSH] OR "end of life care "[tiab] "terminally ill"[MeSH] OR "advanced illness"[tiab] OR " terminal illness"[tiab] OR "end of life"[tiab]
- 3 (Instrument[tiab] OR instruments[tiab] OR measure[tiab] OR measures[tiab] OR questionnaire[tiab] OR questionnaires[tiab] OR scale[tiab] OR scales[tiab] OR tool[tiab] OR tools[tiab] OR survey [tiab] OR test [tiab])
- 4 Instrumentation[sh] OR methods[sh] OR "Validation Studies"[pt] OR "Comparative Study"[pt] OR "psychometrics"[MeSH] OR psychometr*[tiab] OR clinimetr*[tw] OR clinometr*[tw] OR "outcome assessment (health care)"[MeSH] OR "outcome assessment"[tiab] OR "outcome measure*" [tw] OR "observer variation"[MeSH] OR "observer variation"[tiab] OR "Health Status Indicators"[Mesh] OR "reproducibility of results"[MeSH] OR reproducib*[tiab] OR "discriminant analysis"[MeSH] OR reliab*[tiab] OR unreliab*[tiab] OR valid*[tiab] OR coefficient[tiab] OR homogeneity[tiab] OR homogeneous[tiab] OR "internal consistency"[tiab] OR (cronbach*[tiab] AND (alpha[tiab] OR alphas[tiab])) OR (item[tiab] AND (correlation*[tiab] OR selection*[tiab] OR reduction*[tiab])) OR agreement[tiab] OR precision[tiab] OR imprecision[tiab] OR "precise values"[tiab] OR test-retest[tiab] OR (test[tiab] AND retest[tiab]) OR (reliab* [tiab] AND (test[tiab] OR retest[tiab])) OR stability[tiab] OR interrater[tiab] OR inter-rater[tiab] OR intrarater[tiab] OR intra-rater[tiab] OR intertester[tiab] OR inter-tester[tiab] OR intratester[tiab] OR intra-tester[tiab] OR interobserver[tiab] OR inter-observer[tiab] OR intraobserver[tiab] OR intraobserver[tiab] OR intertechnician[tiab] OR inter-technician[tiab] OR intratechnician[tiab] OR intra-technician[tiab] OR interexaminer[tiab] OR inter-examiner[tiab] OR intraexaminer[tiab] OR intra-examiner[tiab] OR interassay[tiab] OR inter-assay[tiab] OR intraassay[tiab] OR intra-assay[tiab] OR interindividual[tiab] OR inter-individual[tiab] OR intraindividual[tiab] OR intra-individual[tiab] OR interparticipant[tiab] OR inter-participant[tiab] OR intraparticipant[tiab] OR intra-participant[tiab] OR kappa[tiab] OR kappa's[tiab] OR kappas[tiab] OR repeatab*[tiab] OR ((replicab*[tiab] OR repeated[tiab]) AND (measure[tiab] OR measures[tiab] OR findings[tiab] OR result[tiab] OR results[tiab] OR test[tiab] OR tests[tiab])) OR generaliza*[tiab] OR generalisa*[tiab] OR concordance[tiab] OR (intraclass[tiab] AND correlation*[tiab]) OR discriminative[tiab] OR "known group"[tiab] OR "factor analysis"[tiab] OR "factor analyses"[tiab] OR dimension*[tiab] OR subscale*[tiab] OR (multitrait[tiab] AND scaling[tiab] AND (analysis[tiab] OR analyses[tiab])) OR "item discriminant"[tiab] OR "interscale correlation*" [tiab] OR error[tiab] OR errors[tiab] OR "individual variability"[tiab] OR (variability[tiab] AND (analysis[tiab] OR values[tiab])) OR (uncertainty[tiab] AND (measurement[tiab] OR measuring[tiab])) OR "standard error of measurement"[tiab] OR sensitiv*[tiab] OR responsive*[tiab] OR ((minimal[tiab] OR minimally[tiab] OR clinical[tiab] OR clinically[tiab]) AND (important[tiab] OR significant[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR (small*[tiab] AND (real[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR "meaningful change"[tiab] OR "ceiling effect"[tiab] OR "floor effect"[tiab] OR "Item response model"[tiab] OR IRT[tiab] OR Rasch[tiab] OR "Differential item functioning"[tiab] OR DIF[tiab] OR "computer adaptive testing"[tiab] OR "item bank"[tiab] OR "cross-cultural equivalence"[tiab]
- 5 #1 AND #2 AND #3 AND #4
- 6 "Protocol"[Publication Type] OR "addresses"[Publication Type] OR "biography"[Publication Type] OR "case reports"[Publication Type] OR "comment"[Publication Type] OR "editorial"[Publication Type] OR "congresses"[Publication Type] OR "consensus development conference"[Publication Type] OR "consensus development conference"[Publication Type] OR "practice guideline"[Publication Type]
- 7 #5 NOT #6
- 8 FILTER: Language (English and Spanish)

Estrategia de búsqueda para Web of Science (WoS)

- 1 TI=("Fatigue" OR "Tiredness" OR "Weakness" OR "Cancer related fatigue")
- 2 TS=(Palliative care OR terminal care OR end of life care OR terminally ill OR advanced illness OR terminal illness OR end of life)
- 3 TS=(Instrument OR instruments OR measure OR measures OR questionnaire OR questionnaires OR scale OR scales OR tool OR tools OR survey OR test)
- 4 TS=("Instrumentation" OR methods" OR "Validation Studies" OR "Comparative Study" OR "psychometrics" OR psychometr*OR clinimetr* OR clinometr* OR "outcome assessment (health care)" OR "outcome assessment" OR "outcome measure* OR "observer variation" OR "observer variation" OR "Health Status Indicators" OR "reproducibility of results" OR reproducib* OR "discriminant analysis" OR reliab* OR unreliab* OR valid* OR coefficient OR homogeneity OR homogeneous OR "internal consistency" OR (cronbach* AND (alpha OR alphas)) OR (item AND (correlation* OR selection* OR reduction*)) OR agreement OR precision OR imprecision OR "precise values" OR test-retest OR (test AND retest) OR (reliab* AND (test OR retest)) OR stability OR interrater OR inter-rater OR intrarater OR intra-rater OR intertester OR inter-tester OR intratester OR intra-tester OR interobserver OR inter-observer OR intraobserver OR intraobserver OR intertechnician OR inter-technician OR intratechnician OR intra-technician OR interexaminer OR inter-examiner OR intraexaminer OR intra-examiner OR interassay OR inter-assay OR intraassay OR intra-assay OR interindividual OR inter-individual OR intraindividual OR intra-individual OR interparticipant OR inter-participant OR inraparticipant OR intra-participant OR kappa OR kappa's OR kappas OR repeatab* OR ((replicab* OR repeated) AND (measure OR measures OR findings OR result OR results OR test OR tests)) OR generaliza* OR generalisa* OR concordance OR (intraclass AND correlation*) OR discriminative OR "known group" OR "factor analysis" OR "factor analyses" OR dimension* OR subscale* OR (multitrait AND scaling AND (analysis OR analyses)) OR "item discriminant" OR "interscale correlation" OR error OR errors OR "individual variability" OR (variability AND (analysis OR values)) OR (uncertainty AND (measurement OR measuring)) OR "standard error of measurement" OR sensitiv* OR responsive* OR ((minimal OR minimally OR clinical OR clinically) AND (important OR significant OR detectable) AND (change OR difference)) OR (small*AND (real OR detectable) AND (change OR difference)) OR "meaningful change" OR "ceiling effect" OR "floor effect" OR "Item response model" OR IRT OR Rasch OR "Differential item functioning" OR DIF OR "computer adaptive testing" OR "item bank" OR "cross-cultural equivalence")
- 5 #1 AND #2 AND #3 AND #4
- 6 ("Protocol" OR "addresses" OR "biography" OR "case reports" OR "comment" OR "editorial" OR "congresses" OR "consensus development conference" OR "consensus development conference" OR "practice guideline")
- 7 #5 NOT #6
- 8 FILTER: Language (English and Spanish)

Estrategia de búsqueda para Open Grey

- 1 ("Fatigue" OR "Tiredness" OR "Weakness" OR "Cancer related fatigue")
- 2 (Palliative care OR terminal care OR end of life care OR terminally ill OR advanced illness OR terminal illness OR end of life)
- 3 (Instrument OR instruments OR measure OR measures OR questionnaire OR questionnaires OR scale OR scales OR tool OR tools OR survey OR test)
- 4 ("Instrumentation" OR methods" OR "Validation Studies" OR "Comparative Study" OR "psychometrics" OR psychometr* OR clinimetr* OR clinometr* OR "outcome assessment (health care)" OR "outcome assessment" OR "outcome measure" OR "observer variation" OR "observer variation" OR "Health Status Indicators" OR "reproducibility of results" OR reproducib* OR "discriminant analysis" OR reliab* OR unreliab* OR valid* OR coefficient OR homogeneity OR homogeneous OR "internal consistency" OR (cronbach* AND (alpha OR alphas)) OR (item AND (correlation* OR selection* OR reduction*)) OR agreement OR precision OR imprecision OR "precise values" OR test-retest OR (test AND retest) OR (reliab* AND (test OR retest)) OR stability OR interrater OR inter-rater OR intrarater OR intra-rater OR intertester OR inter-tester OR intratester OR intra-tester OR interobserver OR inter-observer OR intraobserver OR intraobserver OR intertechnician OR inter-technician OR intratechnician OR intra-technician OR interexaminer OR inter-examiner OR intraexaminer OR intra-examiner OR interassay OR inter-assay OR intraassay OR intra-assay OR interindividual OR inter-individual OR intraindividual OR intra-individual OR interparticipant OR inter-participant OR intraparticipant OR intra-participant OR kappa OR kappa's OR kappas OR repeatab* OR ((replicab* OR repeated) AND (measure OR measures OR findings OR result OR results OR test OR tests)) OR generaliza* OR generalisa* OR concordance OR (intraclass AND correlation*) OR discriminative OR "known group" OR "factor analysis" OR "factor analyses" OR dimension* OR subscale* OR (multitrait AND scaling AND (analysis OR analyses)) OR "item discriminant" OR "interscale correlation" OR error OR errors OR "individual variability" OR (variability AND (analysis OR values)) OR (uncertainty AND (measurement OR measuring)) OR "standard error of measurement" OR sensitiv* OR responsive* OR ((minimal OR minimally OR clinical OR clinically) AND (important OR significant OR detectable) AND (change OR difference)) OR (small* AND (real OR detectable) AND (change OR difference)) OR "meaningful change" OR "ceiling effect" OR "floor effect" OR "Item response model" OR IRT OR Rasch OR "Differential item functioning" OR DIF OR "computer adaptive testing" OR "item bank" OR "cross-cultural equivalence")
- 5 #1 AND #2 AND #3 AND #4
- 6 ("Protocol" OR "addresses" OR "biography" OR "case reports" OR "comment" OR "editorial" OR "congresses" OR "consensus development conference" OR "consensus development conference" OR "practice guideline")
- 7 #5 NOT #6
- 8 FILTER: Language (English and Spanish)

Estrategia de búsqueda para Scielo

- 1 ("Fatiga" OR "Cansancio" OR "Debilidad" OR "Fatiga relacionada con el cáncer")
- 2 (Cuidados paliativos OR atención terminal OR atención al final de la vida OR enfermedad terminal OR enfermedad avanzada OR final de la vida)
- 3 (Instrumento OR instrumentos OR medida OR medidas OR cuestionario OR cuestionarios OR escala OR escalas OR herramienta OR herramientas OR encuesta OR prueba)
- 4 ("Instrumentación" OR métodos" OR "Estudios de validación" OR "Estudio comparativo" OR "psicometría" OR psicometr* OR clinimetr* OR clinometr* OR "evaluación de resultados (atención médica)" OR "evaluación de resultados" OR "medida de resultado*" OR "Variación del observador" OR "variación del observador" OR "Indicadores del estado de salud" OR "reproducibilidad de los resultados" OR reproduc* OR "análisis discriminante" OR confiable* OR no confiable* OR válido* OR coeficiente OR homogeneidad OR homogéneo OR "consistencia interna" OR (cronbach* AND (alfa OR alfa)) OR (elemento AND (correlación* OR selección* OR reducción*)) OR acuerdo OR precisión OR imprecisión OR "valores precisos" OR prueba-reprueba OR (prueba AND reprueba) OR (confiabilidad* AND (prueba OR reprueba)) OR estabilidad OR inter-evaluador OR inter-evaluador OR intra-evaluador OR intra-evaluador OR inter-evaluador OR inter-evaluador OR intra-evaluador OR intra-evaluador OR interobservador OR interobservador OR intraobservador OR intraobservador OR intertécnica OR inter-técnico OR Intratechnician OR intra-técnico OR interexaminador OR inter-examinador OR intra-examinador OR intra-examinador OR interensayo OR interensayo OR intraensayo OR intraensayo OR interindividual OR interindividual OR intraindividual OR intraindividual OR interparticipante OR interparticipante OR intraparticipante OR intraparticipante OR kappa OR kappa OR kappas OR repeat* OR ((replicab* OR repetido) AND (medida OR mide OR hallazgos OR resultado OR resultados OR prueba OR pruebas)) OR generaliza* OR generaliza* OR concordancia OR (intraclase AND correlación*) OR discriminativo OR "grupo conocido" OR "análisis factorial" OR "análisis factorial" OR dimensión* OR subescala* OR (multi-retrato AND escala AND (análisis OR análisis)) OR "discriminante de ítem" OR "correlación entre escalas" OR error OR errores OR "variabilidad individual" OR (variabilidad AND (análisis OR valores)) OR (incertidumbre AND (medición OR medición)) OR "error estándar de medición" OR sensible* OR sensible* OR ((mínimo OR mínimamente OR clínico OR clínicamente) AND (importante OR significativo OR detectable) AND (cambio OR diferencia)) OR (pequeño* AND (real OR detectable) AND (cambio OR diferencia)) OR "cambio significativo" OR "efecto techo" OR "efecto piso" OR "Modelo de respuesta del artículo" OR IRT O Rasch OR "Funcionamiento diferencial de elementos" OR DIF OR "pruebas de adaptación por computadora" OR "banco de elementos" OR "equivalencia transcultural")
- 5 #1 AND #2 AND #3 AND #4
- 6 ("Protocolo" OR "direcciones" OR "biografía" OR "estudio de casos" OR "comentarios" OR "editorial" OR "congresos" OR "conferencia" OR "guía de práctica clínica")
- 7 #5 NOT #6
- 8 FILTER: Language (English and Spanish)

Estrategia de búsqueda para Cochrane

- 1 (Fatigue*:ti,ab OR Tiredness*:ti,ab OR Weakness*:ti,ab OR Cancer related fatigue*:ti,ab)
- 2 (Palliative care*:ti,ab OR terminal care*:ti,ab OR end of life care*:ti,ab OR terminally ill*:ti,ab OR advanced illness*:ti,ab OR terminal illness*:ti,ab OR end of life*:ti,ab)
- 3 (Instrument:ti,ab OR instruments:ti,ab OR measure:ti,ab OR measures:ti,ab OR questionnaire:ti,ab OR questionnaires:ti,ab OR scale:ti,ab OR scales:ti,ab OR tool:ti,ab OR tools:ti,ab OR survey:ti,ab OR test:ti,ab)
- 4 AB ("Instrumentation" OR methods" OR "Validation Studies" OR "Comparative Study" OR "psychometrics" OR psychometr*OR clinimetr* OR clinometr* OR "outcome assessment (health care)" OR "outcome assessment" OR "outcome measure" OR "observer variation" OR "observer variation" OR "Health Status Indicators" OR "reproducibility of results" OR reproducib* OR "discriminant analysis" OR reliab* OR unreliab* OR valid* OR coefficient OR homogeneity OR homogeneous OR "internal consistency" OR (cronbach* AND (alpha OR alphas)) OR (item AND (correlation* OR selection* OR reduction*)) OR agreement OR precision OR imprecision OR "precise values" OR test-retest OR (test AND retest) OR (reliab* AND (test OR retest)) OR stability OR interrater OR inter-rater OR intrarater OR intra-rater OR intertester OR inter-tester OR intratester OR intra-tester OR interobserver OR inter-observer OR intraobserver OR intraobserver OR intertechnician OR inter-technician OR intratechnician OR intra-technician OR interexaminer OR inter-examiner OR intraexaminer OR intra-examiner OR interassay OR inter-assay OR intraassay OR intra-assay OR interindividual OR inter-individual OR intraindividual OR intra-individual OR interparticipant OR inter-participant OR intraparticipant OR intra-participant OR kappa OR kappa's OR kappas OR repeatab* OR ((replicab* OR repeated) AND (measure OR measures OR findings OR result OR results OR test OR tests)) OR generaliza* OR generalisa* OR concordance OR (intraclass AND correlation*) OR discriminative OR "known group" OR "factor analysis" OR "factor analyses" OR dimension* OR subscale* OR (multitrait AND scaling AND (analysis OR analyses)) OR "item discriminant" OR "interscale correlation" OR error OR errors OR "individual variability" OR (variability AND (analysis OR values)) OR (uncertainty AND (measurement OR measuring)) OR "standard error of measurement" OR sensitiv* OR responsive* OR ((minimal OR minimally OR clinical OR clinically) AND (important OR significant OR detectable) AND (change OR difference)) OR (small*AND (real OR detectable) AND (change OR difference)) OR "meaningful change" OR "ceiling effect" OR "floor effect" OR "Item response model" OR IRT OR Rasch OR "Differential item functioning" OR DIF OR "computer adaptive testing" OR "item bank" OR "cross-cultural equivalence")
- 5 #1 AND #2 AND #3 AND #4
- 6 ("Protocol" OR "addresses" OR "biography" OR "case reports" OR "comment" OR "editorial" OR "congresses" OR "consensus development conference" OR "consensus development conference" OR "practice guideline")
- 7 #5 NOT #6
- 8 FILTER: Language (English and Spanish)

Estrategia de búsqueda para CINAHL

- 1 (MH "Fatigue+") OR (MH "Tiredness+") OR (MH "Weakness+")OR (MH "Cancer related fatigue+")
- 2 (MH "Palliative care") OR (MH "terminal care") OR (MH "end of life care ") (MH "terminally ill") OR (MH "advanced illness") OR (MH " terminal illness") OR (MH "end of life")
- 3 TI (Instrument* OR instruments* OR measure* OR measures* OR questionnaire* OR questionnaires* OR scale* OR scales* OR tool* OR tools* OR survey* OR test*)
- 4 (MH "Instrumentación") OR (MH "métodos") OR (MH "Estudios de validación") OR (MH "Estudio comparativo") OR (MH "psicometría") OR psicometr* OR clinimetr* OR clinometr* OR (MH "evaluación de resultados") OR (MH "evaluación de resultados") OR (MH "medida de resultado") OR (MH "Variación del observador") OR (MH "variación del observador") OR (MH "Indicadores del estado de salud") OR (MH "reproducibilidad de los resultados") OR reproduc* OR (MH "análisis discriminante") OR confiable* OR no confiable* OR válido* OR coeficiente* OR homogeneidad* OR homogéneo* OR (MH "consistencia interna") OR cronbach* AND alfa OR alfa* OR (NH "elemento") AND (correlación* OR selección* OR reducción*) OR acuerdo* OR precisión* OR imprecisión* OR (NH "valores precisos") OR prueba-reprueba* OR confiabilidad* OR estabilidad* OR inter-evaluador* OR inter-evaluador* OR intra-evaluador* OR intra-evaluador* OR inter-evaluador* OR inter-evaluador* OR intra-evaluador* OR intra-evaluador* OR interobservador* OR interobservador* OR intraobservador* OR intraobservador* OR intertecnica* OR intertecnico* OR Intratechnician* OR intra-técnico* OR interexaminer* OR inter-examinador* OR intra-examinador* OR intra-examinador* OR interensayo* OR interensayo* OR intraensayo* OR intraensayo* OR interindividual* OR interindividual* OR intraindividual* OR intraindividual* OR interparticipante* OR interparticipante* OR intraparticipante* OR intraparticipante* OR kappa* OR kappa* OR kappas* OR repeatab* OR medida* OR mide* OR hallazgos* OR resultado* OR resultados* OR prueba* OR pruebas* OR generaliza* OR generaliza* OR concordancia* OR (MH "intraclase AND correlación") OR discriminativo* OR (NH "grupo conocido") OR (NH "análisis factorial") OR (MH "análisis factorial") OR dimensión* OR subescala* OR (NH "multi-retrato AND escala") OR (NH "discriminante de ítem") OR (NH "correlación entre escalas ") OR error* OR errores* OR (NH "variabilidad individual ") OR análisis OR valores* OR (NH "error estándar de medición") OR sensible* OR sensible * OR mínimo* OR mínimamente* OR clínico* OR clínicamente* OR pequeño* OR (NH "cambio significativo") OR (NH "efecto techo") OR (NH "efecto piso") OR (NH "Modelo de respuesta del artículo") OR IRT* O Rasch* OR (NH "Funcionamiento diferencial de elementos") OR DIF* OR (NH "pruebas de adaptación por computadora") OR (NH "banco de elementos") OR (NH "equivalencia transcultural")
- 5 #1 AND #2 AND #3 AND #4
- 6 ("Protocol" OR "addresses" OR "biography" OR "case reports" OR "comment" OR "editorial" OR "congresses" OR "consensus development conference" OR "consensus development conference" OR "practice guideline")
- 7 #5 NOT #6
- 8 FILTER: Language (English and Spanish)

Estrategia de búsqueda para EMBASE

- 1 (fatigue/exp OR fatigue* OR tiredness/exp OR tiredness* OR weakness/exp OR weakness* OR cancer related fatigue/exp OR cancer related fatigue*)
- 2 (Palliative care/exp OR terminal care/exp OR end of life care/exp OR terminally ill/exp OR advanced illness/exp OR terminal illness/exp OR end of life/exp)
- 3 (Instrument* OR instruments* OR measure* OR measures* OR questionnaire* OR questionnaires* OR scale* OR scales* OR tool* OR tools* OR survey* OR test*)
- 4 (Instrumentation*:de,ab,ti OR methods* OR validation Studies/exp OR Comparative Study/exp OR psychometrics/exp OR psychometr*:de,ab,ti OR clinimetr* OR clinometr* OR outcome assessment/exp OR outcome assessment* OR outcome measure* OR observer variation/de OR observer variation/de OR Health Status Indicators/de OR reproducibility of results/exp OR reproducib* OR discriminant analysis* OR reliab*: de,ab,ti OR unreliab* OR valid* OR coefficient* OR homogeneity* OR homogeneous* OR internal consistency* OR cronbach*: de,ab,ti OR alpha/exp OR alphas* OR correlation* OR selection* OR reduction* OR agreement* OR precision* OR imprecision* OR precise values* OR test-retest*: de,ab,ti OR reliab* OR stability* OR interrater* OR inter-rater* OR intrarater* OR intra-rater* OR intertester *OR inter-tester* OR intratester* OR intra-tester* OR interobserver* OR inter-observer* OR intraobserver* OR intertechnician* OR inter-technician* OR intratechnician* OR intra-technician* OR interexaminer* OR inter-examiner* OR intraexaminer* OR intra-examiner* OR interassay* OR inter-assay* OR intraassay* OR intra-assay* OR interindividual* OR inter-individual* OR intraindividual* OR intra-individual* OR interparticipant* OR inter-participant* OR intraparticipant* OR intra-participant* OR kappa*:de,ab,ti OR kappa's* OR kappas* OR repeatab* OR replicab* OR repeated* OR generaliza* OR generalisa* OR concordance* OR intraclass*: de,ab,ti OR correlation*: de,ab,ti OR discriminative* OR known group* OR factor analysis* OR dimension* OR subscale* OR multitrait* OR analysis* OR analyses* OR item discriminant* OR interscale correlation* OR error* OR errors* OR individual variability* OR analysis* OR values* OR standard error of measurement* OR sensitiv* OR responsive* OR minimal* OR minimally* OR clinical* OR clinically* OR important* OR significant* OR detectable* OR change* OR difference* OR small* OR real* OR detectable* OR change* OR difference* OR meaningful change* OR ceiling effect* OR floor effect* OR Item response model* OR IRT* OR Rasch*: de,ab,ti OR Differential item functioning* OR DIF*: de,ab,ti OR computer adaptive testing* OR item bank* OR cross-cultural equivalence*)
- 5 #1 AND #2 AND #3 AND #4
- 6 ("Protocol" OR "addresses" OR "biography" OR "case reports" OR "comment" OR "editorial" OR "congresses" OR "consensus development conference" OR "consensus development conference" OR "practice guideline")
- 7 #5 NOT #6
- 8 FILTER: Language (English and Spanish)

2.3 Criterios de inclusión y exclusión

Los estudios elegibles fueron aquellos que cumplieron con los siguientes criterios de inclusión: a) estudios de validación de instrumentos que evalúan FRC en CP, b) evalúan al menos una propiedad de medición, y c) publicados en inglés o español.

Se excluyeron biografías, alocuciones, editoriales, informes de casos, comentarios y congresos, conferencias, guías de práctica y artículos no publicados en idioma inglés o español.

Los títulos y resúmenes fueron seleccionados de forma independiente por 2 revisores. Se recuperó el artículo completo cuando no se pudieron tomar decisiones sobre el resumen y el título por sí solos. Las discrepancias en la selección de los artículos se resolvieron mediante consenso con un tercer revisor.

2.4 Evaluación de la calidad

2 revisores valoraron de manera independiente los artículos a texto completo utilizando la Strobe Checklist para evaluar la calidad de los estudios incluidos (tabla 11). Posteriormente, se evaluó la calidad metodológica para cada estudio usando la lista de verificación de cuatro puntos COSMIN para revisiones sistemáticas y los desacuerdos fueron resueltos por un tercer revisor (21).

Esta herramienta considera 9 propiedades de medición (cajas). Cada cuadro proporciona una puntuación sobre calidad metodológica para las siguientes propiedades de medición: validez de contenido, validez estructural, consistencia interna, validez transcultural, error de medición, fiabilidad, validez de criterio, prueba de hipótesis y sensibilidad al cambio. Cada cuadro contiene de 3 a 35 ítems y cada ítem tiene cuatro opciones de respuesta: "muy bueno", "adecuado", "dudoso" o "inadecuado". La puntuación de cada propiedad psicométrica se obtuvo tomando la puntuación más baja de cualquier elemento para cada propiedad (21).

2.5 Extracción de datos

2 investigadores participaron de forma independiente en el proceso de extracción de datos y cualquier desacuerdo fue resuelto por un tercer revisor. Se extrajo la siguiente información: autor del estudio y fecha de publicación, población de estudio/tipo de cáncer, ámbito, descripción del instrumento, propiedades de medición, puntuación COSMIN y resultados estadísticos de las propiedades de medición.

2.6 Síntesis de datos

La síntesis de datos comprendió 3 pasos. En primer lugar, la calidad metodológica de cada estudio que cumplió con los criterios de inclusión se puntuó utilizando las calificaciones de calidad de la lista de verificación de riesgo de sesgo de COSMIN como "muy buena", "adecuada", "dudosa" o "inadecuada" (21). En segundo lugar, el resultado de cada estudio se calificó según los criterios actualizados para las buenas propiedades de medición de Terwee y Prinsen como "suficiente" (+), "insuficiente" (-) o "indeterminado" (?) (17,22). En tercer lugar, se resumió la evidencia y se calificó la calidad de la evidencia para cada propiedad de medición como "alta", "moderada", "baja" o "muy baja", empleando la calificación de recomendaciones, valoración, desarrollo y evaluaciones del método GRADE (23). 2 investigadores participaron de forma independiente en el proceso de extracción de datos y cualquier desacuerdo fue resuelto por un tercer revisor.

Tabla 11. Calidad de los estudios incluidos (Strobe Checklist).

Autor (es)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	Total
Alawneh et al. (2015)	•	•	•	•	•	•	-	-	-	•	•	-	-	•	•	-	-	•	-	•	•	-	13
Aoun et al. (2011)	•	•	•	•	•	•	-	•	-	-	•	•	-	•	•	-	•	•	-	•	•	•	16
Arraras et al. (2013)	•	•	•	-	•	•	-	•	-	•	•	-	•	•	-	•	•	-	•	•	-	15	
Baró et al. (2011)	•	•	•	•	•	•	-	•	-	-	•	•	-	•	•	-	•	•	-	•	•	•	16
Blazeby et al. (2009)	•	•	•	-	•	•	-	•	-	-	-	-	-	•	•	-	-	•	-	•	•	•	12
Carvajal et al. (2011)	•	•	•	-	•	•	-	•	-	•	•	-	-	•	•	•	-	•	-	•	•	•	15
Carvajal et al. (2013)	•	•	•	•	•	•	-	•	-	-	•	-	-	•	•	-	-	•	-	•	•	•	14
Chinda et al. (2011)	•	•	•	-	•	•	-	•	•	-	•	-	-	•	•	-	-	•	-	•	•	•	14
Claessens et al. (2011)	•	•	•	•	•	•	-	•	-	-	•	-	-	•	•	-	-	•	-	•	•	•	14
Davis et al. (2013)	-	•	•	•	•	•	-	-	-	-	-	•	-	-	-	-	-	•	•	-	•	-	10
Ewing et al. (2004)	•	•	•	•	•	•	-	•	-	-	•	-	-	-	•	-	-	•	•	•	-	•	13
Godoy Fernández et al. (1999)	-	•	•	-	-	-	-	•	-	-	•	-	-	-	•	-	-	•	-	•	•	-	8
Golčić et al. (2018)	•	•	•	-	•	•	-	•	-	•	•	-	-	•	•	-	-	•	•	•	-	•	14
Gretarsdóttir et al. (2016)	•	•	•	-	•	•	-	•	-	•	•	-	-	•	•	-	-	•	-	•	•	•	14
Groenvold et al. (2006)	•	•	•	-	-	•	-	•	-	-	•	-	-	•	•	-	-	•	-	•	-	•	11
Hagelin et al. (2007)	•	•	•	-	-	•	-	•	-	•	•	-	-	•	•	-	-	•	-	•	-	-	11
Hannon et al. (2015)	•	•	•	•	•	•	-	•	-	-	•	-	-	•	•	•	-	•	•	•	-	•	15
Johnsen et al. (2011)	•	•	•	•	-	•	-	•	-	-	•	-	-	•	•	-	-	•	-	-	-	•	11
Kaasa et al. (1995)	•	•	•	-	•	•	-	•	-	-	•	•	-	•	•	-	•	•	-	-	•	-	14
Kim et al. (2006)	-	•	•	-	•	•	-	•	-	-	•	•	-	•	•	•	•	•	-	•	-	•	14
Kim et al. (2009)	•	•	•	-	•	•	-	•	-	-	•	-	-	•	•	-	-	•	-	•	-	-	11
Knobel et al. (2003)	•	•	•	-	•	•	-	•	-	-	•	•	-	•	•	-	•	•	-	•	-	•	15
Kwon et al. (2013)	•	•	•	-	•	•	-	•	-	•	•	-	-	•	•	•	•	•	-	•	-	•	16
Leppert et al. (2012)	•	•	•	-	•	•	-	•	-	-	•	-	-	•	•	-	-	•	-	•	-	•	12
Lyons et al. (2009)	•	•	•	-	•	•	-	•	-	-	•	•	-	•	•	-	-	•	-	•	-	-	11
McMillan et al. (1994)	•	•	•	-	•	•	-	•	•	-	•	-	-	•	•	-	-	-	•	•	-	-	12
Miyashita et al. (2015)	•	•	•	•	•	•	-	•	-	-	•	-	-	•	•	-	-	•	-	•	•	•	14
Miyazaki et al. (2011)	•	•	•	-	•	•	-	•	-	-	•	-	•	•	•	-	-	•	•	•	-	•	14
Mori et al. (2018)	•	•	•	-	•	•	-	•	-	-	•	-	-	•	•	-	-	•	-	•	-	•	12
Mystakidou et al. (2001)	•	•	•	-	•	•	-	•	-	-	•	•	-	•	•	-	•	•	-	•	-	-	13

Autor (es)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	Total
Mystakidou et al. (2004)	-	●	●	-	●	●	-	●	-	-	●	●	-	●	●	-	●	●	-	-	-	-	11
Mystakidou et al. (2008)	●	●	●	-	●	●	-	●	-	-	●	-	-	●	●	-	-	-	-	●	-	-	10
Nekolaichuk et al. (2017)	●	●	●	-	●	●	-	●	-	-	●	-	-	●	●	-	-	●	-	●	●	●	13
Nicklasson et al. (2007)	●	●	●	-	●	●	-	●	-	-	●	-	-	●	●	-	-	-	-	●	●	●	12
Nunes et al. (2014)	●	●	●	-	●	●	-	●	-	-	●	-	-	●	●	-	-	●	-	●	-	-	11
Okuyama et al. (2000)	●	●	●	-	●	●	-	●	-	-	●	-	-	●	●	-	-	●	●	●	-	●	13
Okuyama et al. (2003)	●	●	●	-	●	●	-	●	-	-	●	-	-	●	●	-	-	●	-	●	●	●	13
Osse et al. (2004)	-	●	●	●	●	●	-	●	●	-	●	-	-	●	●	-	-	●	-	●	-	●	13
Osse et al. (2007)	-	●	●	●	●	●	-	●	-	-	●	-	-	●	●	-	-	●	●	-	-	-	11
Ozcelik et al. (2016)	●	●	●	-	●	●	-	●	-	●	●	-	-	●	●	-	-	●	-	●	●	●	14
Paiva et al. (2015)	●	●	●	-	●	●	-	●	-	-	●	-	-	●	●	●	-	-	-	●	●	●	13
Paradowska et al. (2017)	●	●	●	-	●	●	-	●	-	●	●	●	-	●	●	-	●	●	-	●	-	●	15
Pautex et al. (2017)	●	●	●	-	●	●	-	●	-	●	●	-	-	●	●	-	-	●	-	●	-	●	14
Philip et al. (1998)	-	●	●	-	●	●	-	-	-	-	-	-	-	-	●	-	-	-	-	-	-	-	6
Raman et al. (2016)	●	●	●	-	●	●	-	●	-	-	●	●	-	●	●	-	●	-	-	●	●	●	14
Rojas-Concha et al. (2020)	●	●	●	-	●	●	-	●	-	-	●	-	-	●	●	-	-	●	-	●	●	●	13
Shin et al. (2011)	●	●	●	●	●	●	-	●	-	-	●	-	-	●	●	-	-	●	-	●	●	●	14
Shinall et al. (2018)	●	●	●	-	●	●	-	●	-	-	●	-	-	●	●	-	-	-	-	-	-	●	10
Siegert et al. (2014)	●	●	●	-	●	●	-	●	●	-	●	-	-	●	●	-	-	●	●	●	●	●	15
Sprague et al. (2018)	●	●	●	-	●	●	-	●	●	-	●	-	-	●	●	-	-	●	-	●	-	●	13
Sterkenburg et al. (1996)	●	●	●	-	●	●	-	●	-	-	●	●	-	●	●	-	●	●	-	-	-	●	13
Suarez Del Real et al. (2010)	●	●	●	-	●	●	-	●	-	●	●	-	-	●	●	●	-	●	-	●	●	-	14
Tricou et al. (2019)	●	●	●	-	●	●	-	●	-	●	●	-	-	●	●	-	-	●	-	●	●	●	14
Wang et al. (2019)	●	●	●	-	●	●	-	●	-	●	●	-	-	●	●	-	-	●	-	●	-	●	13
Weis et al. (2017)	●	●	●	●	●	●	-	●	●	-	●	-	-	●	●	●	-	●	-	●	-	●	15
Yokomichi et al. (2015)	●	●	●	-	●	●	-	●	-	-	●	-	-	●	●	●	-	●	●	●	●	●	15
Zhang et al. (2015)	●	●	●	-	●	●	-	●	-	-	●	-	-	●	●	-	-	●	●	●	●	●	14

●: Yes

-: No

3. RESULTADOS

Se identificaron un total de 5598 artículos en las diferentes bases de datos. Después de eliminar los artículos duplicados, se examinaron 2966 artículos por título y resumen. De estos, 2488 artículos fueron excluidos después de esta evaluación. Posteriormente, se evaluaron 123 artículos a texto completo. Un total de 66 estudios fueron excluidos por las siguientes razones: no incluyeron una población de estudio en CP (18 artículos), el instrumento no evaluó la fatiga (26 artículos), al menos una propiedad de medición no fue evaluada (16 artículos), el idioma no era inglés ni español (4 artículos) y se trataba de una biografía, dirección, editorial, caso clínico, congreso o conferencia (2 artículos). Los 57 estudios restantes identificados a partir de la búsqueda en las diferentes bases de datos se incluyeron en esta revisión sistemática.

La calidad de los estudios se evaluó con la escala Strobe (tabla 11). Todos los estudios obtuvieron una puntuación superior a 10 puntos, excepto Philip et al. (32) (6 puntos), en el que se evaluó el Edmonton Symptom Assessment System (ESAS), y Godoy Fernández et al. (26) (8 puntos), en el que se evaluó el European Organization for Research and Treatment of Cancer Core Quality of Life Questionnaire (EORTC QLQ-C30). El diagrama de flujo se muestra en la figura 17. Se identificaron 19 instrumentos diseñados específicamente para evaluar la fatiga en personas con cáncer en CP (tabla 12). La mayoría de estos estudios se realizaron en el ámbito hospitalario y en hospicios, mientras que el resto se localizaron a nivel domiciliario, ambulatorio y comunitario (tabla 13).

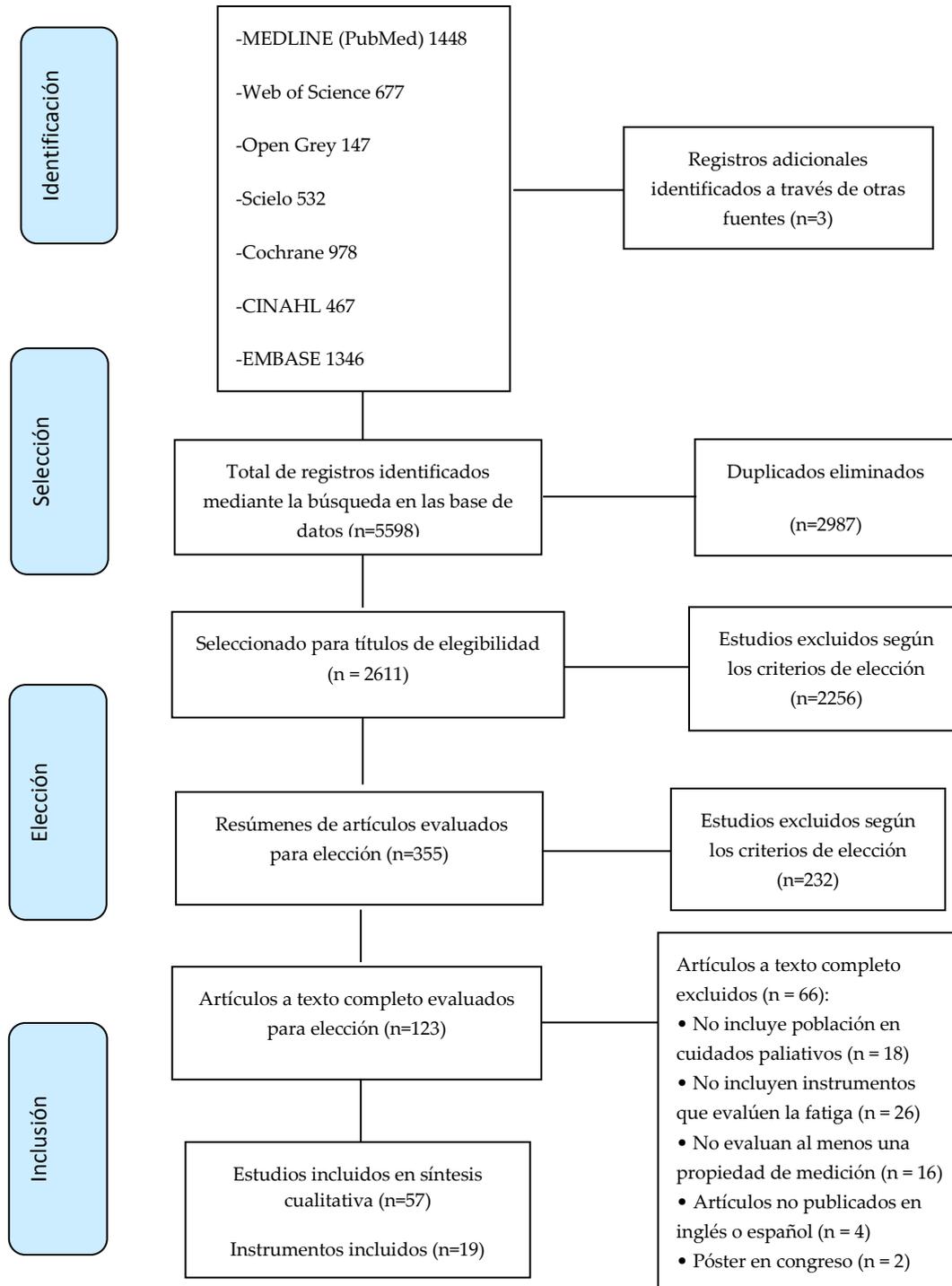


Figura 17. Diagrama de flujo de la selección del estudio.

Tabla 12. Instrumentos identificados para evaluar la fatiga en personas con cáncer en cuidados paliativos.

Instrumentos	Autor (es) / año
Brief Fatigue Index (BFI)	Okuyama et al. (2003), Mystakidou et al. (2008), Davis et al. (2013)
Cambridge Palliative Assessment Schedule (CAMPAS-R)	Ewing et al. (2004)
Cancer Fatigue Scale (CFS)	Okuyama et al. (2000)
European Organization for Research and Treatment of Cancer Quality of Life 15 items Questionnaire for Palliative Care (EORTC QLQ-C15-PAL)	Groenvold et al. (2006), Suarez Del Real et al. (2010), Miyazaki et al. (2011), Shin et al. (2011), Leppert et al. (2012), Arraras et al. (2013), Nunes et al. (2014), Alawneh et al. (2015), Miyashita et al. (2015), Zhang et al. (2015), Ozcelik et al. (2016), Raman et al. (2016), Golčić et al. (2018), Rojas-Concha et al. (2020)
European Organization for Research and Treatment of Cancer Core Quality of Life Questionnaire (EORTC QLQ- C30)	Kaasa et al. (1995), Godoy Fernández et al. (1999), Mystakidou et al. (2001), Knobel et al. (2003), Nicklasson et al. (2007)
European Organization for Research and Treatment of Cancer Quality of Life 12 items Questionnaire (EORTC QLQ-FA12)	Weis et al. (2017)
European Organization for Research and Treatment of Cancer Colorectal Liver Metastases (EORTC QLQ-LMC21)	Blazeby et al. (2009), Paradowska et al. (2017)
Edmonton Symptom Assessment System (ESAS)	Philip et al. (1998), Carvajal et al. (2011, 2013), Chinda et al. (2011), Claessens et al. (2011), Kwon et al. (2013), Hannon et al. (2015), Paiva et al. (2015), Yokomichi et al. (2015), Gretarsdottir et al. (2016), Nikolaichuk et al. (2017), Pautex et al. (2017), Mori et al. (2018), Sprague et al. (2018), Tricou et al. (2019)

Instrumentos	Autor (es) / año
Functional Assessment of Chronic Illness Therapy-Palliative Care (FACIT-PAL)	Lyons et al. (2009), Siegert et al. (2014) , Shinall et al. (2018)
Four-Item Fatigue Scale (FIFS)	Davis et al. (2013)
Hospice Quality of Life Index (HQLI)	McMillan et al. (1994)
Hospice Quality of Life Scale (HQLS)	Kim et al. (2009)
Multidimensional Fatigue Inventory (MFI-20)	Hagelin et al. (2007)
McMaster Quality of Life Scale (MQLS)	Sterkenburg et al. (1996), Kim et al. (2006)
Problems and Needs in Palliative Care Questionnaire (PNPC)	Osse et al. (2004, 2007), Wang et al. (2019)
Perform Questionnaire (PQ)	Baró et al. (2011)
Palliative Care Quality of Life Instrument (PQLI)	Mystakidou et al. (2004)
Symptom Assessment Scale (SAS)	Aoun et al. (2011)
Three-Levels-of-Needs Questionnaire (3LNQ)	Johnsen et al. (2011)

Tabla 13. Características de los estudios y descripción de las medidas de fatiga relacionada con el cáncer en los instrumentos de cuidados paliativos.

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Hospice Quality of Life Index (HQLI)						
McMillan et al. (1994)	68 pacientes terminales y sus cuidadores principales. No especifica los tipos de cáncer presentes en el estudio.	Estados Unidos: Los pacientes fueron reclutados en un hospicio en Florida.	El Índice de calidad de vida de hospicio (HQLI) es una escala de 25 ítems. Incluye bienestar físico, bienestar psicológico / espiritual, bienestar social y bienestar financiero. Cada ítem de satisfacción es una escala analógica visual (EVA) con anclajes adjetivos en cada extremo y una línea de 100 mm entre los anclajes en la que el paciente debe hacer una única marca. Las puntuaciones ponderadas resultantes pueden oscilar entre 0 y 300, donde cero representa la peor calidad de vida posible y 300 representa la mejor calidad de vida posible.	1-Validez de contenido 2-Prueba de hipótesis 3-Sensibilidad al cambio	1-Inadecuada 2-Dudosa 3-Dudosa	- La correlación paciente y cuidador fue muy débil ($r = 0.06$, $p < 0.05$). La correlación entre las estimaciones del paciente y el cuidador del dolor en su peor momento fue moderada ($r = 0.42$, $p < 0.005$). Las correlaciones de Pearson fueron más altas para los elementos relacionados con el contacto físico con otros, qué tan bien come el paciente y qué tan bien duerme ($r = 0.47$, $r = 0.44$, $r = 0.42$, respectivamente). Todas las demás correlaciones fueron muy débiles o cercanas a cero.

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
(EORTC QLQ-C30)						
Kaasa et al. (1995)	247 pacientes tratados con radioterapia paliativa. La mayoría de los pacientes tenían un diagnóstico primario de cáncer de pulmón (32%), cáncer de próstata (13%) o cáncer de mama (23%). Otro diagnóstico incluido; mieloma (4%), cáncer gastrointestinal (5%) y cáncer de recto (4%).	Noruega: Los pacientes fueron reclutados en cuatro hospitales diferentes, en Oslo.	El Cuestionario EORTC QLQ-C30 (versión 1.0) se compone de 9 escalas de elementos múltiples: 5 escalas de funcionamiento, una escala de calidad de vida global (Global QoL) y 3 escalas de síntomas (fatiga, dolor y náuseas / vómitos). Los ítems se puntúan en una escala categórica de 4 puntos que va desde 1 "nada" a 4 "mucho". Para las cinco escalas de funcionamiento y la escala Global QoL, una puntuación más alta representa un mejor nivel de funcionamiento. Para las escalas de síntomas y los ítems, una puntuación alta corresponde a un nivel más alto de sintomatología.	1-Consistencia interna 2-Prueba de hipótesis 3-Sensibilidad al cambio	1-Muy buena 2-Adecuada 3-Adecuada	- α Cronbach \Rightarrow 0.70 (excepto para las escalas de funcionamiento; rol y cognitivo). - Las correlaciones más fuertes se encontraron entre las escalas de funcionamiento físico, funcionamiento de roles ($r = 0.72$, $r = 0.67$) y fatiga ($r = -0.60$, $r = -0.63$). La escala de funcionamiento emocional tuvo una alta correlación con el GHQ-20, tanto antes como después del tratamiento ($r = -0.62$ y $r = -0.71$). Tanto antes como después del tratamiento, la escala de dolor EORTC QLQ-C30 se correlacionó en gran medida con los ítems que evaluaban la intensidad del dolor ($r = 0.85$, $r = 0.78$), la frecuencia del dolor ($r = -0.69$, $r = -0.66$) y la intensidad del dolor medida en una escala analógica visual (escala EVA) ($r = 0.79$, $r = 0.71$).

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
(EORTC QLQ-C30)						
Godoy Fernández et al. (1999)	73 pacientes con cáncer avanzado. No especifica los tipos de cáncer presentes en el estudio.	España: Para la realización del estudio se utilizaron sujetos oncológicos de dos Unidades de Cuidados Paliativos; 45 de ellos atendidos por la Unidad de Cuidados Paliativos del Complejo Hospitalario Torrecárdenas y 28 pacientes que correspondieron al Equipo Coordinado de Cuidados Paliativos de Granada.	El EORTC QLQ-C30 (versión 2.0) es un cuestionario de 30 ítems. Tiene 5 subescalas funcionales, 3 subescalas sintomáticas y una subescala de calidad de vida global. El resto de ítems evalúan individualmente los síntomas adicionales que padecen los pacientes oncológicos. También incluye la subescala de calidad de vida global, que comprende 2 ítems; una se refiere a la salud general y la otra a la calidad de vida en general, enmarcando sus respuestas en una escala entre 1 y 7 puntos. Los elementos restantes están formateados en el tipo Likert de 4 puntos.	1-Consistencia interna 2-Fiabilidad 3-Prueba de hipótesis 4-Sensibilidad al cambio	1-Muy buena 2-Inadecuada 3-Muy buena 4-Muy buena	- α Cronbach \Rightarrow 0.70 (excepto la subescalas de función emocional y función cognitiva, que fueron 0.62 y 0.67 respectivamente). - CCI = > 0.70 (excepto las subescalas de síntomas: náuseas /vómitos, que fue 0.63) - Las correlaciones más altas se encontraron entre la subescala de función de rol y función física ($r = 0.76$). Como era de esperar, también se encontraron correlaciones estadísticamente significativas entre la subescala de función física, la subescala de función de rol y la subescala de fatiga sintomática; $r = 0.76$ y $r = 0.61$, respectivamente. La subescala QL se correlacionó significativamente con la mayoría de las otras subescalas.

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
(EORTC QLQ-C30)						
Mystakidou et al. (2001)	120 pacientes con cáncer avanzado en tratamiento paliativo. Los diagnósticos más prevalentes fueron cáncer de pulmón (25%), cáncer de mama (13.3%), páncreas (13.3%), cervical (10%) y ovárico (8.3%).	Grecia: La muestra se reclutó de la clínica ambulatoria de una unidad de cuidados paliativos en Atenas, Hellas.	La versión griega del EORTC QLQ-C30 (versión 3.0) consta de 30 ítems, incluye 5 escalas funcionales, 3 escalas de síntomas y una de estado de salud global, 5 ítems individuales que evalúan síntomas adicionales y el impacto financiero. Todas las escalas y medidas de un solo elemento varían en puntuación de 0 a 100. Una puntuación alta para una escala funcional representa un nivel de funcionamiento alto / saludable, una puntuación alta para la salud / calidad de vida global representa una alta calidad de vida, y una puntuación alta para una escala / ítem de síntomas representa un alto nivel de síntomas / problemas.	1-Validez estructural 2-Consistencia interna 3-Validez transcultural 4-Prueba de hipótesis 5-Sensibilidad al cambio	1-Dudosa 2-Muy buena 3-Dudosa 4-Adecuada 5-Adecuada	- El EFA resultó en una solución de 6 factores. - α Cronbach \Rightarrow 0.70 (excepto la escala de funcionamiento físico y la escala de síntomas fatiga, náuseas / vómitos y dolor, que fueron $<$ 0.70). - Las correlaciones más fuertes antes del tratamiento se observaron entre el funcionamiento físico, el funcionamiento social y la fatiga; mientras que durante el tratamiento se observaron entre el dolor, el funcionamiento físico, el funcionamiento de roles y las escalas de fatiga. También se observaron correlaciones sustanciales ($r \geq 0.40$), pretratamiento y sobretratamiento entre la escala de funcionamiento físico, el funcionamiento emocional, el funcionamiento social y el ítem de calidad de vida.

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
(EORTC QLQ-C30)						
Knobel et al. (2003)	238 pacientes con cáncer metastásico avanzado tratados con radioterapia paliativa; cáncer de mama (29%), próstata (37%), pulmón (12%), mielomatosis (9%) y otros tipos de cánceres (12%) y 128 pacientes con leucemia y linfoma maligno tratados curativamente con trasplante de células madre y alta dosis de quimioterapia.	Noruega: Los pacientes fueron reclutados para participar en un ensayo clínico aleatorizado.	EORTC QLQ-C30 (versión 3.0) Como anteriormente	1-Validez estructural 2-Prueba de hipótesis al cambio 3-Sensibilidad	1-Adecuada 2-Muy buena 3-Muy buena	- Se realizó un EFA que obtuvo 2 factores. Todos los ítems de fatiga física cargados en el factor uno, "la fatiga física", al igual que todos los ítems de la escala de fatiga (excepto en los pacientes de cuidados paliativos al comienzo del estudio) y todos los ítems de fatiga mental cargados en el segundo factor "la fatiga mental". - El análisis demostró que la escala de fatiga se correlacionó más alto con la escala de fatiga física ($r = 0.67-0.75$; $p > 0.05$) en comparación con la escala de fatiga mental ($r = 0.49-0.61$; $p > 0.05$). Las correlaciones ítem-escala entre los ítems de la escala de fatiga y la escala de fatiga física fueron consistentemente más altas que entre los ítems de la escala de fatiga y la escala de fatiga mental.

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
(EORTC QLQ-C30)						
Nicklasson et al. (2007)	112 pacientes con cáncer de pulmón avanzado o mesotelioma pleural, no susceptibles de tratamiento curativo o que prolongue la vida.	Suecia: Departamento de Medicina Respiratoria, Sahlgrenska Hospital Universitario.	La versión sueca de la EORTC QLQ-C30 (versión 2.0) contiene 30 ítems que cubren problemas de salud relevantes para una amplia gama de pacientes con cáncer, 24 de los cuales se agregan en escalas de múltiples ítems que miden el funcionamiento, los síntomas (fatiga, dolor y náuseas / vómitos), salud global y calidad de vida. Los 6 ítems restantes evalúan síntomas adicionales asociados con el cáncer y el impacto económico percibido de la enfermedad y el tratamiento.	1-Consistencia interna 2-Prueba de hipótesis 3-Sensibilidad al cambio	1-Muy buena 2-Muy buena 3-Muy buena	- α Cronbach \Rightarrow 0.70 (excepto la escala de funcionamiento cognitivo) - Se demostró una fuerte correlación entre el funcionamiento emocional y la escala de ansiedad HADS, que a su vez se correlacionó con la Global QoL. La escala de depresión HADS se correlacionó con todas las escalas funcionales, fatiga y pérdida de apetito. La subescala de intensidad BPI (BPI-I) se correlacionó fuertemente ($r = 0.72$) con la escala de dolor EORTCQLQ-C30, y principalmente ($r = > 0.4$; $p < 0.05$) con los ítems de dolor LC13. La subescala de función BPI (BPI-F) se correlacionó con todas las escalas de funcionamiento y la calidad de vida global, así como con las medidas de disnea y dolor, tanto en la EORTC QLQ-C30 como en la LC13.

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
McMaster Quality of Life Scale (MQLS)						
Sterkenburg et al. (1996)	262 participantes; 83 pacientes de cuidados paliativos, 60 familiares y 119 miembros del personal. La localización del cáncer primario fue en; pulmón (21%), mama (17%), intestino (16%), próstata (14%), páncreas (4%), vejiga (4%), linfoma (3%) y otros (20%).	Canadá: Hamilton y London, Ontario. En Hamilton participaron cuatro centros; Hospitales Chedoke McMaster, St. Joseph's, St. Peter's y la Clínica SIS. En London, los pacientes que participaron fueron ubicados en una unidad de cuidados paliativos para pacientes hospitalizados en el hospital Parkwood.	La escala de calidad de vida de McMaster (MQLS) abarca cuatro dimensiones: física, emocional, social y espiritual. Sus 32 ítems incluyen 11 ítems físicos y 20 ítems no físicos; estado funcional (2 ítems), funcionamiento social (2 ítems), estado emocional (5 ítems), cognición (2 ítems), sueño y descanso (2 ítems), energía y vitalidad (3 ítems), rol (4 ítems) y un ítem de calidad de vida general (el cuestionario se puntúa de 1 a 7, donde una puntuación más alta significa una calidad de vida más alta).	1-Consistencia interna 2-Fiabilidad 3-Prueba de hipótesis 4-Sensibilidad al cambio	1-Dudosa 2-Inadecuada 3-Muy buena 4-Muy buena	- α Cronbach = ≥ 0.80 para la escala general; paciente (0.80), familia (0.87) y personal (0.89), excepto todas sus subescalas; física, emoción, cognición, energía, función, descanso, rol y social, que fueron < 0.70 . - ICC = confiabilidad intra-evaluador (0.63-0.95) y confiabilidad entre evaluadores (Inicial; 0.51- 0.64, Time 2; 0.51-0.61). - Correlaciones moderadas entre el índice de calidad de vida de Spitzer y el MQLS ($r = 0.70$; $p < 0.05$). La correlación entre el índice de Spitzer puntuado por el personal y el MQLS puntuado por el paciente fue moderada ($r = 0.50$; $p < 0.05$).

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
(MQLS)						
Kim et al. (2006)	70 pacientes adultos con cáncer terminal; 70 familiares y 28 enfermeras. Tipos de cáncer; cáncer del aparato digestivo (57.4%), aparato respiratorio (19.1%) y otros cánceres (23.5%).	Corea: Se seleccionaron siete centros de cuidados paliativos y hospicios clínicos coreanos de entre 20 centros distribuidos por Corea por el método de muestreo por conglomerados.	La versión coreana de la escala de calidad de vida McMaster (K-MQLS) Como anteriormente	1-Validez de contenido 2-Consistencia interna 3-Validez transcultural 4-Fiabilidad 5-Prueba de hipótesis 6-Sensibilidad al cambio	1-Inadecuada 2-Muy buena 3-Inadecuada 4-Inadecuada 5-Adecuada 6-Adecuada	- α Cronbach = 0.81 en pacientes, 0.74 en familias y 0.91 en enfermeras (excepto la subescala física en enfermeras). - CCI = En términos de confiabilidad intra evaluador, fueron los siguientes: paciente a paciente 0.93, familia a familia 0.76 y enfermera a enfermera 0.94. El coeficiente de fiabilidad test-retest de paciente a paciente fue de 0.63 después de una semana. - El coeficiente de correlación de Pearson para la validez concurrente entre la puntuación K-MQLS de los pacientes y la puntuación QLI de los médicos mostró un nivel de correlación moderado y fue estadísticamente significativo ($r = 0.55$; $p < 0.01$). Este estudio encontró que el ranking de correlaciones positivas fue enfermera ($r = 0.94$; $p < 0.05$), paciente ($r = 0.93$; $p < 0.05$), familia ($r = 0.76$; $p < 0.05$), paciente y enfermera ($r = 0.66$; $p < 0.05$).

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Edmonton Symptom Assessment System (ESAS)						
Philip et al. (1998)	80 sujetos con cáncer incurable. No especifica los tipos de cáncer presentes en el estudio.	Australia: Servicio de Cuidados Paliativos de Sydney Central (el 50% eran pacientes hospitalizados en la unidad de cuidados paliativos, el 25% pacientes ambulatorios y el 25% pacientes de atención domiciliaria).	El sistema (ESAS) se modificó ligeramente para la población australiana. La ESAS consta de escalas analógicas visuales que miden el dolor y otros ocho síntomas que se encuentran comúnmente en los cuidados paliativos. Los síntomas incluidos son (10): dolor, debilidad, náuseas, depresión, ansiedad, somnolencia, posible apetito, sensación de bienestar, dificultad para respirar y una pregunta adicional; "En las últimas 24 h, ¿cuánto alivio han proporcionado los tratamientos para el dolor o los medicamentos?" La gravedad de cada síntoma se califica de 0 a 10 (0 = ausencia de síntoma, 10 = peor intensidad posible).	1-Validez transcultural 2-Prueba de hipótesis 3-Sensibilidad al cambio	1-Inadecuada 2-Adecuada 3-Adecuada	- Kappa ponderado con RSCL = 0.45-0.58; y con BPI = 0.58-0.61.

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
(ESAS)						
Carvajal et al. (2011)	171 pacientes con cáncer avanzado. El 47.9% de los pacientes habían sido diagnosticados con tumores gastrointestinales, de mama (13%), ginecológicos (10%), pulmonares (10.9%), hematológicos (6%) y otros (8.5%).	España: Servicios de Oncología y Hematología del Hospital de la Universidad de Navarra, Pamplona.	La versión en español (ESAS) tiene once escalas de calificación numérica que exploran los síntomas frecuentes en pacientes de cuidados paliativos. Los síntomas incluidos son (10): dolor, fatiga, náuseas, depresión, ansiedad, somnolencia, dificultad para respirar, apetito, dificultad para dormir y sensación de bienestar. Se incluye una escala en blanco para síntomas adicionales. La gravedad de cada síntoma se califica de 0 a 10 (0 = ausencia de síntoma, 10 = peor intensidad posible). Cada escala numérica se interpreta independientemente de las otras escalas, aunque también es posible calcular una SDS total a partir de la suma total de todos los síntomas.	1-Validez de contenido 2-Validez estructural 3-Consistencia interna 4-Validez transcultural 5-Fiabilidad 6-Prueba de hipótesis 7-Sensibilidad al cambio	1-Inadecuada 2-Adecuada 3-Muy buena 4-Adecuada 5-Muy buena 6-Muy buena 7-Muy buena	- Se realizó un EFA, que encontró 3 dominios centrales: "suave", "físico duro" y "emocional". - α Cronbach = 0.75 Correlación de Spearman = 0.65 - 0.94. - La validez concurrente con la RSCL encontró una buena correlación en los síntomas físicos (Kappa hasta 0.66) pero una correlación débil en los síntomas emocionales (Kappa 0.35).

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
(ESAS)						
Chinda et al. (2011)	44 pacientes con cáncer avanzado en hospital. Tipos de cáncer; ginecológico (36.5%), mama (29.3%), gastrointestinal (62.1%), pulmón (15.8%), hematológico (21.3%), cabeza y cuello (15.9%), urinario (6.8%), musculoesquelético (4.5%) y desconocido (2.3%).	Tailandia: Departamentos de Medicina, Cirugía, Ginecología, Ortopedia y Otorrinolaringología del Hospital Ramathibodi.	Versión tailandesa (ESAS-thai) Los síntomas incluidos en esta versión son (9): dolor, fatiga, náuseas, depresión, ansiedad, somnolencia, dificultad para respirar, apetito y bienestar. Esta versión no incluye una escala en blanco para síntomas adicionales. Resto como anteriormente	1-Validez de contenido 2-Consistencia interna 3-Validez transcultural	1-Inadecuada 2-Inadecuada 3-Inadecuada	- α Cronbach = 0.75 (entorno hospitalario) y 0.89 (entorno ambulatorio), a excepción de los síntomas "fatiga" y "somnolencia", que fueron 0.69 para ambos.
Claessens et al. (2011)	Tres muestras de conveniencia; la primera muestra 8 expertos en cuidados paliativos, la segunda muestra 4 pacientes, 5 familiares y 5 enfermeras, y la última 23 pacientes. Localización del tumor primario; riñón (4%) pulmón (17%) mama (13%) colon (9%) estómago (4%) útero (4%) tiroides (4%) próstata (13%) páncreas (22%) bilis (4%) y laringe (4%).	Bélgica: Pacientes con cáncer ingresados en las unidades de cuidados paliativos de los hospitales universitarios de Lovaina.	M-ESAS. Los síntomas incluidos en esta versión son (12): dolor, fatiga, náuseas, depresión, ansiedad, somnolencia, disminución del apetito, malestar / sensación de tristeza, dificultad para respirar, obstrucción intestinal, sequedad de boca y trastornos del sueño. Esta versión no incluye una escala en blanco para síntomas adicionales.	1-Validez de contenido 2-Validez transcultural 3-Prueba de hipótesis 4-Sensibilidad al cambio	1-Inadecuada 2-Inadecuada 3-Adecuada 4-Adecuada	- Solo para los síntomas de ansiedad hubo una fuerte correlación entre la angustia ($r = 0.78$; $p < 0.05$) y el tiempo de aparición del síntoma. En el 25% de los síntomas, las puntuaciones de angustia fueron más altas que las puntuaciones de aparición (dolor, náuseas y disnea). Para todos los demás síntomas, las puntuaciones de aparición del síntoma fueron más altas que la angustia.

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
(ESAS)						
Carvajal et al. (2013)	66 pacientes con cáncer avanzado. Diagnóstico de cáncer primario; cáncer gastrointestinal (36%), genitourinario (24%), pulmón (5%), mama (18%) y otros (17%).	España y Guatemala: Pacientes de los departamentos de oncología del hospital universitario de Navarra (España) e Instituto Nacional del Cáncer (Guatemala).	Versión en español (ESAS-r). Los síntomas incluidos son (10): dolor, cansancio, náuseas, depresión, ansiedad, somnolencia, dificultad para respirar, falta de apetito, insomnio y bienestar. Se incluye una escala en blanco para síntomas adicionales. Resto como anteriormente	1-Consistencia interna 2-Validez transcultural 3-Fiabilidad 4-Prueba de hipótesis 5-Sensibilidad al cambio	1-Muy buena 2-Inadecuada 3-Muy buena 4-Muy buena 5-Muy buena	- α Cronbach= 0.86 - ICC = 0.71-0.94. - Se encontró una correlación negativa ($r = -0.38$ a -0.45 ; $p < 0.05$) para 5 ítems (dolor, fatiga, somnolencia, falta de apetito y bienestar). Sin embargo, no se halló una correlación significativa con PPS para los otros síntomas (náuseas, ansiedad, depresión, dificultad para dormir y respirar).
Kwon et al. (2013)	163 pacientes con cáncer avanzado; enfermedad localmente avanzada, en recaída o refractaria o metastásica. Tipos de cáncer: tracto gastrointestinal (37%), malignidad torácica (17%), cabeza / cuello (11%), hepatobiliar (10%), genitourinario / ginecológico (9%), hematológico (7%), mama (4%) y otros (5%).	Nueve departamentos de oncología de hospitales universitarios o de un hospital de veteranos.	Versión coreana (K-ESAS). Los síntomas incluidos son (10): dolor, fatiga, náuseas, depresión, ansiedad, somnolencia, pérdida de apetito, sensación de bienestar, disnea y alteración del sueño. Esta versión evalúa la intensidad media de los síntomas experimentados durante las 24 horas anteriores. Esta versión no incluye una escala en blanco para síntomas adicionales.	1-Validez de contenido 2-Consistencia interna 3- Validez transcultural 4-Fiabilidad 5- Prueba de hipótesis 6-Sensibilidad al cambio	1-Inadecuada 2-Inadecuada 3-Adecuada 4-Muy buena 5- Muy buena 6- Muy buena	- α Cronbach = 0.88 - Los coeficientes de correlación de Pearson entre la línea de base y después de dos a cuatro horas variaron de $r = 0.72$ ($p < 0.05$) a $r = 0.87$ ($p < 0.05$). - Los coeficientes de correlación de Pearson entre las puntuaciones de síntomas de K-ESAS y las puntuaciones de síntomas de K-MDASI variaron de $r = 0.70$ -0.83 ($p < 0.05$). Para el K-HADS, la validez concurrente fue buena para la ansiedad ($r = 0.73$; $p < 0.05$) pero moderada para la depresión ($r = 0.40$; $p < 0.05$).

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
(ESAS)						
Hannon et al. (2015)	202 pacientes ambulatorios con cáncer avanzado. El sitio de tumor más común fue; gastrointestinal (21.8%), pulmón (16.3%), mama (15.8%) y ginecológico (11.9%).	Canadá: Los participantes fueron reclutados en clínicas de cuidados paliativos ambulatorios en el Princess Margaret Cancer Centre en Toronto.	Estreñimiento y sueño del sistema de evaluación de síntomas de Edmonton (ESAS-CS). Se incluyen dos síntomas comunes adicionales de ESAS, estreñimiento y sueño (en lo sucesivo, la escala se denominará Estreñimiento y sueño del sistema de evaluación de síntomas de Edmonton) y llamamos a la suma de los 11 síntomas ESAS-CS. Los síntomas incluidos son (11): dolor, cansancio, náuseas, depresión, ansiedad, somnolencia, dificultad para respirar, apetito, bienestar, sueño y estreñimiento. Esta versión no incluye una escala en blanco para síntomas adicionales. Resto como anteriormente	1-Consistencia interna 2-Fiabilidad 3-Prueba de hipótesis 4-Sensibilidad al cambio	1-Muy buena 2-Dudosa 3-Muy buena 4-Muy buena	- α Cronbach = 0.83 - ICC = 0.69 - Las puntuaciones totales de ESAS-CS y ESAS-r-CS se correlacionaron con el total de MSAS (rho de Spearman 0.62 y 0.64, respectivamente). La correlación de los síntomas individuales con los síntomas de MSAS osciló entre $r = 0.54-0.80$ para ESAS-CS y $r = 0.52-0.74$ para ESAS-r-CS.

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
(ESAS)						
Hannon et al. (2015b)	Como anteriormente	Como anteriormente	La versión revisada del Sistema de Evaluación de Síntomas del Sueño y Estreñimiento de Edmonton (ESAS-r-CS) incluye los mismos nueve elementos que el ESAS, pero proporciona definiciones de síntomas, especifica el período de tiempo como "ahora " y reordena los síntomas en grupos, con exámenes físicos primero, psicológicos segundo y bienestar al final. Para el presente estudio de validación, los ítems "estreñimiento" y "sueño" se incluyen en la ESAS-r, (ESAS-r-CS). Para el elemento "dormir", se establece el intervalo de tiempo en "durante las últimas 24 horas" en lugar de "ahora".	1-Consistencia interna 2-Prueba de hipótesis 3-Sensibilidad al cambio	1-Muy buena 2-Muy buena 3-Muy buena	- α Cronbach = 0.86 - Las puntuaciones totales de ESAS-CS y ESAS-r-CS se correlacionaron con el total de MSAS (rho de Spearman 0.62 y 0.64, respectivamente). La correlación de los síntomas individuales con los síntomas de MSAS osciló entre $r = 0.54-0.80$ para ESAS-CS y $r = 0.52-0.74$ para ESAS-r-CS.

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
(ESAS)						
Paiva et al. (2015)	249 pacientes con cáncer avanzado; enfermedad localmente avanzada, en recaída o refractaria o metastásica. Los tipos de cáncer primario más comunes fueron de mama (34.1%) y colorrectal (17.7%).	Brasil: Los pacientes fueron reclutados en el Barretos Cancer Hospital, Sao Paulo.	Versión brasileña (ESAS). Los síntomas incluidos son (10): dolor, fatiga, náuseas, ansiedad, pérdida de apetito, depresión, somnolencia, sensación de bienestar, disnea y alteraciones del sueño. Esta versión no incluye una escala en blanco para síntomas adicionales. Resto como anteriormente	1-Validez de contenido 2-Consistencia interna 3-Validez transcultural 4-Fiabilidad 5-Prueba de hipótesis 6-Sensibilidad al cambio	1-Inadecuado 2-Muy buena 3-Muy buena 4-Dudosa 5-Muy buena 6-Muy buena	- α Cronbach = 0.861 - ICC = 0.94 - Para la validez convergente, los coeficientes de correlación de Spearman entre las puntuaciones de los síntomas de la ESAS y las puntuaciones de los síntomas de la EORTC QLQ-C30 y la ESS correspondientes oscilaron entre $r = 0.520$ ($p < 0.05$) y $r = 0.814$ ($p < 0.05$).
Yokomichi et al. (2015)	292 pacientes adultos con cáncer en tratamiento en cuidados paliativos. Localización; pulmón (26%), páncreas (13%), colon / recto (9%), hígado (8%), mama (8%), cabeza / cuello (5%), vías biliares (5%), esófago (4%), riñón / vejiga (4%), estómago (3%), ovario / útero (3%), próstata (1%), sangre (1%), tejidos blandos (1%), desconocidos (2%) y otros (7%).	Japón: Hospital del Centro Nacional del Cáncer Este, Hospital Universitario Okayama, Hospital Internacional St. Luke, Centro Médico de la Cruz Roja Japonesa y Hospital General Okayama Saiseikai.	Versión japonesa (ESAS-r). Los síntomas incluidos son (9): dolor, cansancio, somnolencia, náuseas, falta de apetito, dificultad para respirar, depresión, ansiedad y bienestar. Esta versión no incluye una escala en blanco para síntomas adicionales. Resto como anteriormente	1-Validez de contenido 2-Consistencia interna 3-Validez transcultural 4-Fiabilidad 5-Prueba de hipótesis 6-Sensibilidad al cambio	1-Inadecuada 2-Inadecuada 3-Muy buena 4-Dudosa 5-Muy buena 6-Muy buena	- α Cronbach = 0.87 - ICC = 0.90 - Los coeficientes de correlación de Pearson de cada puntuación de síntomas de la ESAS-r con la MDASI y la EORTC QLQ-C30 correspondientes variaron de $r = 0.45-0.80$ ($p < 0.05$).

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
(ESAS)						
Gretarsdottir et al. (2016)	143 pacientes con cáncer avanzado. Diagnóstico de cáncer; mama (24%) hematológico (24%) gastrointestinal (18%) genitourinario (15%) pulmón (6%) cerebro (3%) y otros (7%).	Islandia: Pacientes ambulatorios y hospitalizados que reciben tratamiento en el Hospital Universitario Landspítali.	La versión islandesa (ESAS-r) es una versión modificada de la ESAS original. Los síntomas incluidos son (9): dolor, cansancio, somnolencia, náuseas, falta de apetito, dificultad para respirar, depresión, ansiedad y bienestar. Esta versión no incluye una escala en blanco para síntomas adicionales. Resto como anteriormente	1-Consistencia interna 2-Prueba de hipótesis 3-Sensibilidad al cambio	1-Inadecuada 2-Muy buena 3-Muy buena	- α Cronbach = 0.85 - El análisis de validez mostró correlaciones significativas de moderadas a fuertes entre siete puntuaciones de síntomas coincidentes en la ESAS-r y MDASI, que van desde $r = 0.64$ - 0.86 ($p < 0.05$).
Nekolaichuk et al. (2017)	15 pacientes con enfermedad progresiva incurable (cancerosa y no cancerosa). Diagnóstico de cáncer primario; gastrointestinal (33%), genitourinario (20%), pulmón (13%), otros cánceres (27%) y no cancerosos (7%).	Canadá: La muestra se reclutó en dos centros urbanos de Canadá, a través de las unidades de cuidados paliativos.	Versión francesa (ESAS-r). Los síntomas incluidos son (9): dolor, cansancio, somnolencia, náuseas, falta de apetito, dificultad para respirar, depresión, ansiedad y bienestar. Se incluye una escala en blanco para síntomas adicionales. Resto como anteriormente	1-Validez de contenido 2-Validez transcultural	1-Inadecuada 2-Inadecuada	

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
(ESAS)						
Pautex et al. (2017)	124 pacientes con cáncer avanzado, con síntomas estables (sin modificación de tratamiento en las últimas 24 horas). Las principales enfermedades oncológicas fueron; pulmón (24%), mama (23%) y gastrointestinal (21%).	Suiza y Francia: Los pacientes fueron reclutados en diferentes entornos de cuidados paliativos especializados ambulatorios y hospitalarios en la parte francesa de Suiza y en Francia.	Versión francesa del Sistema de evaluación de síntomas de Edmonton (F-ESAS). Los síntomas incluidos son (9): dolor, fatiga, náuseas, depresión, ansiedad, somnolencia, disnea, apetito y bienestar. Esta versión no incluye una escala en blanco para síntomas adicionales. Resto como anteriormente	1-Validez de contenido 2-Consistencia interna 3-Prueba de hipótesis 4-Sensibilidad al cambio	1-Inadecuada 2-Muy buena 3-Adecuada 4-Adecuada	- α Cronbach = 0.77 - Los ítems más estables fueron el dolor y el apetito (correlación de Spearman = 0.88; $p < 0.05$). Los ítems con las correlaciones más bajas fueron fatiga (correlación de Spearman = 0.67; $p < 0.05$) y disnea (correlación de Spearman = 0.70; $p < 0.05$). La correlación entre las tres versiones de F-ESAS fue fuerte (la correlación de Spearman varió de $r = 0.66-0.95$; $p < 0.01$).

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
(ESAS)						
Mori et al. (2018)	292 pacientes con cáncer avanzado (con y sin ascitis sintomática). Localización del tumor primario; pulmón (26%), hígado / vías biliares / páncreas (25%), esófago, estómago, colon y recto (16%), mama (8%), riñón, vejiga y próstata (5%), cabeza / cuello (5%), ovario / útero (3%) y otros (11%).	Japón: Pacientes que reciben servicios de cuidados paliativos en cinco instituciones.	Edmonton Symptom Assessment System: Ascites Modification (ESAS: AM) es una versión modificada de ESAS con la adición de dos elementos relacionados con la ascitis, "distensión abdominal" y "movilidad". El ESAS: AM Los síntomas incluidos son (11): somnolencia, cansancio, dolor, movilidad, náuseas, distensión abdominal, bienestar, dificultad para respirar, falta de apetito, ansiedad y depresión. Se incluye una escala en blanco para síntomas adicionales. Resto como anteriormente	1-Validez de contenido 2-Validez estructural 3-Consistencia interna 4-Validez transcultural 5-Fiabilidad 6-Prueba de hipótesis 7-Sensibilidad al cambio	1-Inadecuada 2-Adecuada 3-Inadecuada 4-Muy buena 5-Dudosa 6-Muy buena 7-Muy buena	- El EFA identificó 2 factores: "síntomas físicos" (9 ítems; $\alpha = 0.89$) y "síntomas psicológicos" (2 ítems; $\alpha = 0.92$). - α Cronbach = 0.89 - ICC = 0.93 - Correlación muy fuerte con la puntuación total de MDASI ($r = 0.81$, $p < 0.001$); correlación de moderada a fuerte con el estado de salud global / subescalas de calidad de vida y funcionales de la EORTC QLQ-C30 ($r = -0.40$ a -0.61 , $p < 0.01$). Los ítems "distensión abdominal" y "movilidad" mostraron una fuerte correlación con la subescala de ascitis de la EORTC QLQ-PAN26 ($r = 0.70$, $p < 0.001$); y la subescala de funcionamiento físico de EORTC QLQ-C30 ($r = -0.68$, $p < 0.001$), respectivamente. La puntuación total de la ESAS: AM también se correlacionó muy fuertemente con la de la ESAS: AM-24h ($r = 0.90$, $p < 0.001$).

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
(ESAS)						
Mori et al. (2018b)	Como anteriormente	Como anteriormente	Sistema de evaluación de síntomas de Edmonton: Modificación de la ascitis - 24 horas (ESAS: AM-24h) Como anteriormente Centrándose en las 24 horas anteriores, para explorar las diferencias entre los puntajes de los síntomas en diferentes marcos de tiempo; ESAS: AM (actual) VS ESAS: AM-24h (las últimas 24 horas).	1-Validez de contenido 2-Consistencia interna 3-Validez transcultural 4-Prueba de hipótesis 5-Sensibilidad al cambio	1-Inadecuada 2-Inadecuada 3-Muy buena 4-Muy buena 5-Muy buena	- α Cronbach = 0.92 - La puntuación total de la ESAS: AM también se correlacionó muy fuertemente con la de la ESAS: AM-24h ($r = 0.90$, $p < 0.001$).
Sprague et al. (2018)	229 pacientes con diferentes tipos de cáncer avanzado (cáncer de pulmón, colorrectal, gastrointestinal, ginecológico, mama y otros).	Nueva Zelanda: Pacientes inscritos en un servicio de hospicio comunitario.	ESAS. Los síntomas incluidos son (12): dolor, cansancio, náuseas, depresión, ansiedad, somnolencia, apetito, bienestar, dificultad para respirar, estreñimiento, insomnio y complejidad de la atención (finalización del personal). Se incluye una escala en blanco para síntomas adicionales. Resto como anteriormente	1-Validez estructural 2-Prueba de hipótesis 3-Sensibilidad al cambio	1-Muy buena 2-Muy buena 3-Muy buena	- Modelo de Rasch de crédito parcial. El Rasch inicial mostró un PSI de 0.77, lo que indicó una fiabilidad satisfactoria, pero el ajuste general del modelo fue deficiente ($X^2 = 155.41$; $p < 0.01$) y todos los elementos mostraron umbrales desordenados. - Todas las correlaciones residuales excedieron el punto de corte $r = 0.20$. La correlación más alta se encontró entre "depresión y ansiedad", seguida de "bienestar y complejidad" y "somnolencia y apetito".

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
(ESAS)						
Tricou et al. (2019)	45 pacientes diagnosticados de cáncer avanzado, seguidos por el equipo de cuidados paliativos y sin deterioro cognitivo. Tipos de cáncer: mama / ginecológico (29%), cáncer de sangre (20%), próstata / genitourinario (9%), gastrointestinal (18%), cabeza / cuello (4%), pulmón (7%), melanoma (11%) y otros (2%).	Francia: Los pacientes fueron reclutados en la unidad de cuidados paliativos, la clínica para pacientes hospitalizados de cuidados paliativos y la clínica ambulatoria de cuidados paliativos del Hospital Universitario de Lyon Sud.	Versión francesa de la Escala de evaluación de síntomas modificada de Edmonton (ESAS12-F). Los síntomas incluidos son (12): dolor, cansancio, somnolencia, náuseas, vómitos, estreñimiento, apetito, falta de aire, ansiedad, depresión, sueño y bienestar. Esta versión no incluye una escala en blanco para síntomas adicionales. Resto como anteriormente	1-Validez transcultural	1-Inadecuada	

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Cancer Fatigue Scale (CFS)						
Okuyama et al. (2000)	307 pacientes con diferentes tipos de cáncer (30 pacientes de la unidad de cuidados paliativos). Localización del cáncer; mama (50.2%), pulmón (31.9%) y otros (17.9%).	Japón: Los sujetos del presente estudio eran pacientes ambulatorios y hospitalizados con cáncer en el Hospital del Centro Nacional del Cáncer del Este, Tokio.	Cancer Fatigue Scale (CFS) consta de 3 subescalas: aspectos físicos, afectivos y cognitivos de la fatiga y evalúa la naturaleza multidimensional de la fatiga. Cada elemento se califica en una escala de 1 (nada) a 5 (mucho). El rango de respuesta posible para cada puntuación de la subescala es de 0 a 28 para la subescala física y de 0 a 16 para la subescala afectiva y cognitiva (las puntuaciones se ajustan a 0 como un estado sin fatiga). Por lo tanto, la fatiga total se calcula como la suma de estos puntajes de subescala, siendo el puntaje total máximo 60. Cuantos más altos son los puntajes totales, más severa es la fatiga.	1-Validez estructural 2-Consistencia interna 3-Validez transcultural 4-Fiabilidad 5-Prueba de hipótesis 6-Sensibilidad al cambio	1-Inadecuada 2-Muy buena 3-Muy buena 4-Dudosa 5-Muy buena 6-Muy buena	- Se realizó un EFA. Se identificaron 3 factores. La primera subescala constaba de 7 ítems relacionados con las sensaciones físicas de fatiga. La segunda 4 ítems que evaluaban la actividad afectiva y la tercera 4 ítems que evaluaban la atención y la memoria. - α Cronbach = 0.88. - Los coeficientes de correlación promedio test-retest fueron $r = 0.69$; $p < 0.01$. La validez convergente del CFS fue demostrada por las correlaciones de Pearson con los puntajes VAS. Cada factor se correlacionó significativamente con las puntuaciones de la EVA. En referencia a la asociación entre la escala de fatiga por cáncer y el estado funcional, HADS y MMS, la puntuación del estado funcional se asoció significativamente con las puntuaciones de la subescala física y afectiva.

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Brief Fatigue Index (BFI)						
Okuyama et al. (2003)	252 pacientes con cáncer avanzado. Localización del cáncer; gastrointestinal (20%), pulmón (21%), hepatobiliar / páncreas (14%), hematológico (13%), cabeza / cuello (13%), mama (10%) y otros (10%).	Japón: Se realizaron muestreos de pacientes en las clínicas ambulatorias de 6 divisiones de oncología del National Cancer Center Hospital East, Chiba: la Unidad de Cuidados Paliativos, Oncología Torácica, Oncología Gastrointestinal, Oncología de Cabeza y Cuello, Oncología Hepatobiliar y Pancreática y Oncología Quimioterapéutica.	La versión japonesa del Brief Fatigue Inventory (BFI-J) consta de 9 ítems que utilizan una escala numérica del 0 al 10. Los primeros 3 ítems piden a los pacientes que describan su fatiga ahora, en su nivel habitual, y en su peor nivel durante las últimas 24 horas, utilizando los puntos de anclaje "sin fatiga" y "la fatiga más fuerte que puedas imaginar". Los siguientes 6 ítems piden a los pacientes que describan cuánta fatiga ha interferido con diferentes aspectos de su vida durante las 24 horas anteriores. La puntuación general del BFI-J se calcula como el valor medio de estos 9 elementos.	1-Validez estructural 2-Consistencia interna 3-Validez transcultural 4-Prueba de hipótesis 5-Sensibilidad al cambio	1-Dudosa 2-Muy buena 3-Muy buena 4-Muy buena 5-Muy buena	- Se realizó un EFA. Los resultados de la prueba de detección del BFI-J sugirieron una solución de un factor. - α Cronbach = 0.96. - El elemento "habitual", el elemento "peor" y la puntuación global para el BFI-J se correlacionaron significativamente con los de los otros instrumentos de fatiga. Se encontraron correlaciones en los tamaños del efecto entre la puntuación global BFI y las subescalas de fatiga CFS y POMS (coeficientes de correlación de 0.7 o más). De los tres puntajes BFI enumerados anteriormente, el puntaje global tuvo consistentemente los coeficientes de correlación más altos con otros instrumentos de fatiga. Sin embargo, el tamaño del efecto entre la subescala de vigor POMS y los otros instrumentos fue pequeño o moderado (coeficiente de correlación de $r = 0.23-0.42$).

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Brief Fatigue Index (BFI)						
Mystakidou et al. (2008)	102 pacientes que padecen cáncer incurable y reciben cuidados paliativos. Diagnóstico; urogenital (31.4%), pulmón (26.5%), mama (18.6%), gastrointestinal (10.8%) y otros (12.7%).	Grecia: Los pacientes fueron reclutados en una unidad de cuidados paliativos en Atenas.	La versión griega del Inventario Breve de Fatiga (BFI-Gr). Como anteriormente	1-Validez estructural 2-Consistencia interna 3-Validez transcultural 4-Fiabilidad 5-Prueba de hipótesis 6-Sensibilidad al cambio	1-Adecuada 2-Muy buena 3-Dudosa 4-Dudosa 5-Muy buena 6-Muy buena	- Se realizó un EFA. Los resultados para el BFI-Gr sugirieron una solución de un factor. El valor propio fue 6.62 para este factor, seguido de 0.75 y 0.45 para el segundo y tercer factor, respectivamente. - α Cronbach = 0.954. - La confiabilidad test-retest de la escala mostró que el coeficiente de concordancia fue 0.901. - La evaluación de las relaciones entre el BFI-Gr y el resto de medidas mostró correlaciones estadísticamente significativas entre todas las medidas evaluadas ($r = 0.47-0.76$, $P < 0.0005$).
Davis et al. (2013)	65 pacientes hospitalizados o ambulatorios con cáncer avanzado. No especifica los tipos de cáncer presentes en el estudio.	Estados Unidos: Pacientes reclutados en el departamento de medicina paliativa en ambulatorios y hospitales en Cleveland.	El Inventario Breve de Fatiga (BFI). Como anteriormente	1-Prueba de hipótesis 2-Sensibilidad al cambio	1-Adecuada 2-Adecuada	- El FIFS y el BFI se correlacionaron con ciertos ítems de la ICD-10. La correlación de Spearman entre el FIFS y el BFI fue $r = 0.69$ el día 1 y $r = 0.73$ 1 semana después ($p < 0.05$).

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Cambridge Palliative Assessment Schedule (CAMPAS-R)						
Ewing et al. (2004)	109 pacientes adultos que se encontraban en fase paliativa de una enfermedad progresiva, siendo atendidos en su domicilio. No especifica los tipos de cáncer presentes en el estudio.	Reino Unido: los pacientes fueron reclutados con la ayuda de profesionales de los sectores de atención primaria y secundaria.	El Programa de Evaluación Paliativa de Cambridge (CAMPAS-R) se compone de 6 síntomas (dolor, náuseas / vómitos, estreñimiento, diarrea, dificultad para respirar y ansiedad). En esta versión se incluyen 4 elementos más; ansiedad del cuidador, depresión / sensación de depresión, depresión / sensación de depresión del cuidador y por último, un ítem de fatiga. Los síntomas se puntúan, utilizando escalas analógicas visuales (EVA) de 100 mm, en dos dimensiones.	1-Consistencia interna 2-Prueba de hipótesis 3-Sensibilidad al cambio	1-Muy buena 2-Muy buena 3-Muy buena	- α Cronbach = 0.77 (para la gravedad) y 0.80 (para la interferencia). - Las puntuaciones de gravedad del dolor y de interferencia del dolor en CAMPAS-R se correlacionaron con las puntuaciones de gravedad e interferencia correspondientes en BPI y EORTC QLQ-C30. Se encontraron correlaciones fuertes entre los ítems de disnea en EORTC QLQ-C30 y CAMPAS-R y entre ítems de estreñimiento en las dos medidas. Los ítems de fatiga en CAMPAS-R produjeron correlaciones moderadamente buenas con la escala de fatiga EORTC QLQ-C30. Finalmente, hubo correlaciones fuertes entre las puntuaciones de ansiedad del paciente en CAMPAS-R y la subescala de ansiedad HADS que entre las medidas de depresión.

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Palliative Care Quality of Life Instrument (PQLI)						
Mystakidou et al. (2004)	120 pacientes con cáncer en fase terminal que reciben únicamente tratamiento paliativo. Tipos de cáncer; pulmón (27%), gastrointestinal (21%), mama (14%), vía urinaria (12%), próstata (8%), cabeza / cuello (3%), mieloma múltiple (2%), hueso / sarcoma (2%), sarcoma de tejidos blandos (2%), glioblastoma multiforme (2%), astrocitoma (2%), linfoma no Hodgkin (2%), melanoma (2%), útero cervical (2%), vulva (1 %) y vagina (1%).	Grecia: Los pacientes fueron reclutados de la Unidad de Alivio del Dolor y Cuidados Paliativos del Hospital Areteion (Atenas).	El Instrumento de Calidad de Vida en Cuidados Paliativos (PQLI) es un cuestionario de 28 ítems, compuesto por seis escalas de múltiples ítems (2 escalas funcionales, 1 escala de síntomas, 1 escala de elección de tratamiento, 1 escala psicológica) y una escala de un solo ítem (calidad en general de vida).	1-Validez de contenido 2-Validez estructural 3-Consistencia interna 4-Validez transcultural 5-Prueba de hipótesis 6-Sensibilidad al cambio	1-Inadecuada 2-Adecuada 3-Muy buena 4-Dudosa 5-Muy buena 6-Muy buena	- Se realizó un EFA que incluyó 7 factores; factor 1 "actividad", factor 2 "autocuidado", factor 3 "estado de salud", factor 4 "elección de tratamiento", factor 5 "apoyo", factor 6 "comunicación" y factor 7 "afecto psicológico". - α Cronbach = ≥ 0.70 . - Entre el PQLI y el instrumento AQEL, las correlaciones fueron fuertes en todas las escalas, oscilando entre $r = 0.44$ y $r = 0.94$ ($p < 0.05$). Hubo correlaciones significativas entre el EORTC QLQ-C30 y los ítems relevantes del PQLI, que fueron desde $r = 0.79$ - 0.97 ($p < 0.05$), especialmente entre los ítems "dolor" ($r = 0.97$) y "falta de apetito" ($r = 0.97$). También se encontraron correlaciones significativas con las escalas PQLI y las correspondientes escalas AQEL y EORTC QLQ-C30.

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Problems and Needs in Palliative Care questionnaire (PNPC)						
Osse et al. (2004)	64 pacientes que tenían una variedad de diagnósticos de cáncer, estaban en una fase paliativa de tratamiento y vivían en casa. Tumores primarios; mama (45.3%), pulmón (4.7%), esófago / estómago (1.6%), colon / recto (17.2%), ovario (7.8%), melanoma maligno (6.3%), cuello uterino (4.7%), próstata (3.1 %), cuello / cabeza (1.6%), hematológico / sistema linfático (1.6%), cáncer primario de hígado (1.6%), páncreas (1.6%) y origen primario desconocido (6.3%).	Países Bajos: La muestra estuvo constituida por pacientes adultos, no institucionalizados, residentes en sus propios domicilios. Los pacientes fueron reclutados a través de la Universidad de Nijmegen.	La versión holandesa de Cuidados paliativos para pacientes con cáncer (PNPC) tiene 138 elementos y cubre 13 dominios de problemas y necesidades de cuidados paliativos de los pacientes con cáncer avanzado. El PNPC consta de la parte del problema y la parte de la necesidad de cuidados. En cuanto al sistema de "sí" = 2, "algo" / "tanto como ahora" = 1 y "no" = 0). Las puntuaciones más altas indican más problemas y necesidades de atención más fuertes.	1-Validez de contenido 2-Duda 2-Consistencia interna 3-Validez transcultural 4-Prueba de hipótesis 5-Sensibilidad al cambio	1-Inadecuada 2-Dudosa 3-Inadecuada 4-Adecuada 5-Adecuada	- α Cronbach \Rightarrow 0.70 (excepto para "actividades de rol" y "problemas en consultas"). - Casi todas las dimensiones de "problema" en el PNPC mostraron una correlación significativa con las dimensiones relacionadas de HRQL. La dimensión "síntomas físicos" del PNPC se correlacionó con todas las escalas de síntomas físicos de la EORTC QLQ-C30 y con el ítem COOP "quejas físicas". Las dimensiones del PNPC, "cuestiones psicológicas", "actividades de rol" y "cuestiones financieras y administrativas" correlacionaron $r = -0.40$ con los dominios de HRQL correspondientes.

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
(PNPC)						
Osse et al. (2007)	94 pacientes holandeses con una amplia variedad de cánceres; el cáncer de mama fue el tumor más prevalente (39%), seguido del colorrectal (15%), pulmón (8%), próstata (6%), ovario (5%) y melanoma maligno (4%).	Como anteriormente	La versión corta de Cuidados paliativos de pacientes con cáncer (PNPC-sv) tiene 33 ítems y cubre 8 dominios de problemas y necesidades de cuidados paliativos de pacientes con cáncer avanzado. El sistema de puntuación PNPC-sv para la evaluación psicométrica, se adaptaron el método de puntuación del cuestionario original y las recomendaciones del investigador que desarrolló el PNPC-sv ("sí" = 2, "algo" / "tanto como ahora" = 1 y "no" = 0). Los puntajes más altos indican más problemas y necesidades de atención más fuertes. Las propiedades psicométricas de la parte del problema y la parte de la necesidad de cuidados se determinaron por separado.	1-Consistencia interna 2-Validez de criterio 3-Sensibilidad al cambio	1-Inadecuada 2-Inadecuada 3-Inadecuada	- α Cronbach = ≥ 0.70 (excepto "aspecto problemático" de las dimensiones síntomas físicos, problemas sociales y problemas económicos, que fueron 0.65, 0.61 y 0.69 respectivamente). - Todas las dimensiones en la versión corta muestran altas correlaciones con las dimensiones originales del PNPC, muy por encima del requisito de validez de constructo $r = 0.80$. Solo la dimensión de los problemas sociales no se correlacionó tan bien, posiblemente porque esta dimensión no era muy coherente para empezar.

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
(PNPC)						
Wang et al. (2019)	174 pacientes. Los diagnósticos comunes de los pacientes fueron cáncer de pulmón, nasofaringe y colorrectal, y casi el 60% de los pacientes estaban en estadio IV.	China: Posteriormente, los pacientes fueron reclutados a través de una muestra de conveniencia de tres hospitales terciarios; Southwestern Medical University, North Sichuan Medical College y Zunyi Medical University.	La versión en chino mandarín de los cuidados paliativos de los pacientes con cáncer - versión corta (PNPC-sv) Como anteriormente	1-Validez de contenido 2-Validez estructural 3-Consistencia interna 4-Validez transcultural 5-Prueba de hipótesis 6-Sensibilidad al cambio	1-Inadecuada 2-Adecuada 3-Inadecuada 4-Adecuada 5-Muy buena 6-Muy buena	- Se realizó un CFA y se identificaron 2 factores; factor 1 "parte del problema" y factor 2 "parte que necesita atención". El factor de carga varió de 0.12 a 0.79 para la "parte del problema" y de 0.23 a 0.87 para la "parte que necesita atención". - α Cronbach = 0.88 (para la escala total de la parte del problema) y 0.91 (para la parte de necesidad de cuidados). α Cronbach para las subescalas osciló entre 0.58-0.85. - Se detectaron correlaciones estadísticamente significativas ($p < 0.05$) entre el PNPC-sv y el EORTC QLQ-C30 en la mayoría de las subpuntuaciones y puntuaciones totales. Las correlaciones de ítem a total fueron generalmente aceptables, con la mayoría de los valores por encima de $r = 0.40$.

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
(EORTC QLQ-C15-PAL)						
Groenvold et al. (2006)	41 pacientes (cáncer de mama, estómago, próstata, colorrectal y otros) y 66 profesionales de la salud en cuidados paliativos (en cada país, al menos tres profesionales de la salud eran médicos y tres enfermeras).	Dinamarca, Holanda, Noruega, España, Suecia y Reino Unido: Hospicios, departamentos de medicina paliativa, cuidados paliativos domiciliarios y otros.	EORTC QLQ-C15-PAL (versión abreviada de EORTC QLQ-C30) consta de 15 preguntas que incluyen dos escalas funcionales, dos de síntomas de múltiples ítems, dos de síntoma único y cuatro elementos individuales. Utiliza una escala Likert de cuatro puntos: 1 (nada), 2 (un poco), 3 (bastante), y 4 (mucho). Las puntuaciones más altas para la escala funcional y la escala global de salud / calidad de vida indican una mejor calidad de vida. Las escalas de síntomas más altas indican una calidad de vida más baja.	1-Validez de contenido 2-Fiabilidad 3-Validez de criterio 4-Sensibilidad al cambio	1-Inadecuada 2-Inadecuada 3-Inadecuada 4-Inadecuada	- Kappa ponderado = 0.66 – 0.88. - Las correlaciones entre las puntuaciones predichas y observadas variaron de $r = 0.87$ (escala de función cognitiva) a $r = 0.97$ (escala de fatiga) ($p < 0.05$).

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
(EORTC QLQ-C15-PAL)						
Suarez Del Real et al. (2010)	83 pacientes con una variedad de cáncer terminal (gastrointestinal, mama, ginecológico, cabeza / cuello, pulmón y otros).	México: Pacientes que asisten al INCan y la UCP de la Ciudad de México.	La EORTC QLQ-C15-PAL es una abreviatura de 15 ítems de la versión mexicano-español del cuestionario EORTC QLQ-C30, que comprende una escala de estado de salud global, dos escalas funcionales de múltiples ítems, dos escalas de síntomas de varios ítems y cinco escalas de un solo ítem (náuseas y vómitos, disnea, insomnio, pérdida de apetito y estreñimiento). Las respuestas se miden en escalas Likert de 4 a 7 niveles (7 para el estado de salud global y 4 para las escalas restantes), que se transforman linealmente en una puntuación de 0 a 100, donde 100 representa la mejor salud global, el mejor estado funcional y 0 los peores síntomas.	1-Consistencia interna 2-Validez transcultural 3-Fiabilidad 4-Prueba de hipótesis 5-Sensibilidad al cambio	1-Inadecuada 2-Dudosa 3-Inadecuada 4-Muy buena 5-Muy buena	- α Cronbach = > 0.70 (excepto la escala de fatiga, que fue 0.67). - ICC = No especificado - Las mejores puntuaciones en la escala de salud global / calidad de vida se asocian con una mejor función física, menos fatiga, menos dolor y mejor apetito (CC 0.46, 0.43, 0.36, 0.37, respectivamente; $p < 0.001$ para todas las correlaciones). La escala física se correlacionó inversamente con la pérdida de apetito (CC - 0.379; $p < 0.0001$). Cuanto mayor es el KPS, mejor es la HRQL, especialmente en las escalas físicas, de fatiga, disnea e insomnio. El nivel de hemoglobina en sangre se correlacionó pobremente con la HRQL. Cuanto mejor era el nivel de albúmina sérica, mejor era la HRQL, especialmente observada en las escalas; emocional, de dolor y de apetito.

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
(EORTC QLQ-C15-PAL)						
Miyazaki et al. (2011)	32 pacientes con cáncer que reciben cuidados paliativos. Localización del tumor primario; colon, estómago, páncreas, pulmón y otros.	Japón: Todos los pacientes estaban en el departamento de medicina general de un hospital universitario de Tokio.	El EORTC QLQ-C15-PAL es una abreviatura de 15 ítems de la versión japonesa del cuestionario EORTC QLQ-C30. La subescala de función física de 5 ítems en EORTC QLQ-C30 se reduce a 3 ítems relacionados con el entorno de cuidados paliativos en EORTC QLQ-C15-PAL, la subescala de función emocional de 4 ítems se reduce a 2 ítems, la subescala de fatiga de 3 ítems se reduce a 2 ítems, las dos subescalas para el estado de salud global / calidad de vida se reducen a una subescala, y la subescala de náuseas / vómitos de dos preguntas se reduce a una sola pregunta.	1-Consistencia interna 2-Validez transcultural 3-Fiabilidad 4-Validez de criterio 5-Sensibilidad al cambio	1-Inadecuada 2-Dudosa 3-Inadecuada 4-Inadecuada 5-Inadecuada	- α Cronbach \Rightarrow 0.70 (excepto la subescala de fatiga 0.404). - ICC = 0.93 - La correlación ítem-escala para los ítems EORTC QLQ-C15-PAL osciló entre $r = 0.74-0.95$. Los CCI entre las puntuaciones EORTC QLQ-C15-PAL y las correspondientes puntuaciones EORTC QLQ-C30 fueron $r = 0.93$ y superiores ($r = 0.933-0.947$).

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
(EORTC QLQ-C15-PAL)						
Shin et al. (2011)	102 pacientes con cáncer que reciben cuidados paliativos. Tipos de cáncer; cerebro (2%), hígado (21.6%), pulmón (19.6%), hueso / columna (24.5%), peritoneo (20.6%) y otros (21.6%).	Corea: Se realizó una encuesta transversal multicéntrica de unidades cuidados paliativos hospicios.	La versión coreana de la EORTC QLQ-C15-PAL incorpora dos escalas funcionales (física y emocional), dos escalas de síntomas de varios ítems (fatiga, dolor), cinco escalas de síntomas de un solo ítem (náuseas y vómitos, disnea, insomnio, pérdida del apetito y estreñimiento) y una escala de calidad de vida de un solo ítem. Todos los ítems se puntúan en escalas Likert de 4 puntos que van de 1 (nada) a 4 (mucho), con la excepción de un ítem en la escala global de salud / calidad de vida, que utiliza una escala analógica lineal modificada de 7 puntos.	1-Consistencia interna 2-Validez transcultural 3-Prueba de hipótesis 4-Sensibilidad al cambio	1-Muy buena 2-Muy buena 3-Muy buena 4-Muy buena	- α Cronbach = 0.67-0.85. - Correlaciones entre escalas de síntomas; la escala de fatiga se correlacionó moderadamente con la pérdida de apetito ($r = 0.74$; $p < 0.05$), dolor ($r = 0.61$; $p < 0.05$), insomnio ($r = 0.55$; $p < 0.05$) y náuseas y vómitos ($r = 0.48$; $p < 0.05$) escalas. La calidad de vida global se correlacionó moderadamente con la fatiga ($r = -0.43$; $p < 0.05$), pero solo se correlacionó débilmente con el funcionamiento emocional, náuseas y vómitos, insomnio, pérdida de apetito y estreñimiento.

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
(EORTC QLQ-C15-PAL)						
Leppert et al. (2012)	129 pacientes con cáncer avanzado. Tipos de cáncer: pulmón (20.16%), colon (14.73%), riñón (9.30%), próstata (9.30%), mama (6.98%), cabeza / cuello (6.20%), ovario (5.43%), páncreas (5.43%), cuello uterino (3.10%), estómago (3.10%), endometrio (2.33%), vejiga urinaria (1.55%), hígado (1.55%), vulva (1.55%), esófago (1.55%) y otras localizaciones (7.75%).	Polonia: Cátedra y Departamento de Medicina Paliativa de Poznan (unidad de cuidados paliativos, asistencia domiciliaria y guardería).	La versión polaca de la EORTC QLQ-C15-PAL añade dos elementos adicionales (náuseas / vómitos y estreñimiento) para la evaluación de los síntomas (dolor, actividad, náuseas, vómitos, estreñimiento, anorexia, bienestar, disnea, estado de ánimo deprimido, ansiedad y somnolencia).	1-Consistencia interna 2-Validez transcultural 3-Prueba de hipótesis 4-sensibilidad al cambio	1-Dudosa 2-Muy buena 3-Muy buena 4-Muy buena	- Se encontraron valores de α Cronbach tanto para la escala funcional (rango: 0.830-0.925; 0.830-0.932) como para la de síntomas (rango: 0.784-0.940; 0.794-0.941). - La validez concurrente se estableció correlacionando las puntuaciones EORTC QLQ-C15-PAL con las escalas ESAS, BPI-SF y Karnofsky. Los resultados demostraron correlaciones significativas ($p < 0.05$) entre las escalas de síntomas EORTC QLQ-C15-PAL y los ítems individuales y la mayoría de los ítems de la ESAS. También se observaron numerosas correlaciones significativas entre las escalas funcionales de la EORTC QLQ-C15-PAL y los ítems de la ESAS, especialmente en el caso de la global QL.

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
(EORTC QLQ-C15-PAL)						
Arraras et al. (2013)	116 pacientes con cáncer en estadio IV con metástasis óseas.	España: Servicio de Radioterapia, Complejo Hospitalario de Navarra.	La versión en español de la EORTC QLQ-C15-PAL. Este cuestionario incluye cuatro escalas de varios ítems y 6 ítems individuales.	1-Consistencia interna 2-Validez transcultural 3-Prueba de hipótesis 4-Sensibilidad al cambio	1-Muy buena 2-Muy buena 3-Muy buena 4-Muy buena	- α Cronbach \Rightarrow 0.70, excepto fatiga 0.69 (1ª evaluación) y dolor 0.59 (2ª evaluación). - Las correlaciones más altas en la primera evaluación fueron entre fatiga y funcionamiento físico ($r = -0.67$; $p < 0.05$) y entre fatiga y funcionamiento emocional ($r = -0.53$; $p < 0.05$). En la segunda evaluación, las correlaciones más altas también fueron entre fatiga y funcionamiento físico ($r = -0.73$; $p < 0.05$) y entre fatiga y funcionamiento emocional ($r = -0.51$; $p < 0.05$). En ambas evaluaciones, la global QL tuvo correlaciones moderadamente altas con el funcionamiento físico ($r = 0.66$ en la primera evaluación y $r = 0.46$ en la segunda; $p < 0.05$), fatiga ($r = -0.68$ y $r = -0.60$; $p < 0.05$) y dolor ($r = -0.51$ y $r = -0.49$; $p < 0.05$).

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
(EORTC QLQ-C15-PAL)						
Nunes et al. (2014)	104 pacientes ambulatorios con cáncer avanzado. Los tipos de cáncer más frecuentes fueron el de pulmón y el de mama (50.96%).	Brasil: Los pacientes fueron reclutados en el Hospital das Clínicas de la Universidad de Sao Paulo.	La versión brasileña puntúa en una escala de 7 puntos que va de 0 (tan malo como puede ser) a 7 (excelente), estado funcional (funcionamiento físico y emocional) y estado de los síntomas (fatiga, náuseas y vómitos, dolor, disnea, insomnio, pérdida de apetito, estreñimiento) en una escala Likert de 1 (nada) a 4 (mucho).	1-Validez estructural 2-Consistencia interna 3-Validez transcultural 4-Validez de criterio 5-Sensibilidad al cambio	1-Muy buena 2-Muy buena 3-Muy buena 4-Inadecuada 5-Inadecuada	- CFA. Los 3 factores identificados fueron; "Estado de salud global", "escala funcional" y "escala de síntomas". - α Cronbach = ≥ 0.70 (excepto fatiga 0.58 y funcionamiento emocional 0.65). - Se obtuvieron correlaciones positivas y de moderadas a perfectas entre EORTC QLQ-C15-PAL y EORTC QLQ-C30 en un rango de $r = 0.30-1$; Se obtuvieron correlaciones moderadas y fuertes entre EORTC QLQ-C15-PAL, BPI y BDI.
Alawneh et al. (2015)	175 pacientes oncológicos hospitalizados. Tipo primario de cáncer; mama (9.7%), gastrointestinal (25.7%), genitourinario (9.7%), hematológico (19.4%), pulmón (9.7%), sarcoma (8%), cerebro (3.4%), cabeza / cuello (4.6%) y otros (10.3%).	Jordán: Pacientes ingresados en King Hussein Cancer Center, un centro oncológico terciario en Amman.	La versión árabe de la EORTC QLQ-C15-PAL incluye tres escalas; escalas funcionales de varios elementos, escalas de síntomas de varios elementos y una escala de estado de salud global / calidad de vida de un solo elemento.	1-Validez estructural 2-Consistencia interna 3-Validez transcultural 4-Prueba de hipótesis 5-sensibilidad al cambio	1-Adecuada 2-Inadecuada 3-Adecuada 4-Muy buena 5-Muy buena	- El CFA cumplió con los criterios de bondad de ajuste; GFI, CFI, NFI y NNFI (> 0.90). Los 5 factores identificados fueron; factor 1 "emocional", factor 2 "fatiga", factor 3 "dolor", factor 4 "físico" y factor 5 "síntomas". - α Cronbach = 0.898 - Todos los coeficientes de correlación ítem-escala excedieron el valor establecido de $r = 0.40$.

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
(EORTC QLQ-C15-PAL)						
Miyashita et al. (2015)	312 pacientes con cáncer. Localizaciones del tumor primario; colon / recto (30%), mama (23%) y estómago (12%).	Japón: Sala de oncología para pacientes hospitalizados, clínica ambulatoria de oncología del Centro Médico Internacional de la Universidad Médica de Saitama y pacientes en siete unidades de cuidados paliativos.	La versión japonesa consta de 15 preguntas que incluyen dos escalas funcionales, dos escalas de síntomas y cinco escalas de síntomas de un solo elemento. Las opciones de respuesta a estos 14 ítems utilizan una escala Likert de cuatro puntos: 1 (nada), 2 (un poco), 3 (bastante), y 4 (mucho). Un ítem adicional, la escala global de salud / calidad de vida, utiliza una escala numérica de siete puntos del 1 (muy deficiente) al 7 (excelente). Las puntuaciones más altas para la escala funcional y la escala global de salud / calidad de vida indican una mejor calidad de vida. Sin embargo, las escalas de síntomas más altas indican una calidad de vida más baja.	1-Validez estructural 2-Consistencia interna 3-Validez transcultural 4-Fiabilidad 5-Prueba de hipótesis 6-Sensibilidad al cambio	1-Adecuada 2-Inadecuada 3-Muy buena 4-Inadecuada 5-Muy buena 6-Muy buena	- Se realizó un EFA y se estableció el número de factores en nueve según la versión original; "Estado físico", "estado emocional", "fatiga", "dolor", "náuseas / vómitos", "disnea", "insomnio", "pérdida de apetito" y "estreñimiento". - α Cronbach = 0.76 – 0.86 - ICC = 0.52 – 0.77 - Todas las subescalas, especialmente el funcionamiento físico, la fatiga y el dolor, se correlacionaron significativamente con el estado de rendimiento ECOG autoinformado. En las correlaciones entre las subescalas EORTC QLQ-C15-PAL, el funcionamiento físico, el funcionamiento emocional y la calidad de vida se correlacionaron moderadamente positivamente ($r = 0.32-0.47$; $p < 0.05$). Estas tres subescalas se correlacionaron negativamente con las subescalas de síntomas.

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
(EORTC QLQ-C15-PAL)						
Zhang et al. (2015)	243 pacientes de cuidados paliativos con cáncer avanzado. Tipo de cáncer: gástrico (10.7%), pulmón (38.8%), hígado (7.1%), recto (5.1%), mama (13.3%), cervical (3.6%), cabeza / cuello (2.6%), neoplasia cerebral / neoplasia de médula espinal (1%), pancreático (1.5%), colangiocarcinoma (1.5%) y otros (14.8%).	China: Los pacientes fueron reclutados de Tianjin Cancer Hospital	La versión china de la EORTC QLQ-C15-PAL consta de 15 ítems distribuidos entre dos subescalas funcionales (física y emocional), dos escalas de síntomas de múltiples ítems (fatiga, dolor), cinco escalas de síntomas de un solo ítem (náuseas y vómitos, disnea, insomnio, pérdida de apetito y estreñimiento) y un solo ítem en la escala de calidad de vida. Todos los ítems se puntúan en escalas Likert de 4 puntos que van de 1 (Nada) a 4 (Mucho), con la excepción de un ítem en la escala global de salud / calidad de vida, que utiliza una escala analógica lineal de 7 de puntos.	1-Consistencia interna 2-Validez transcultural 3-Validez de criterio 4-Sensibilidad al cambio	1-Inadecuada 2-Muy buena 3-Inadecuada 4-Inadecuada	- α Cronbach = ≥ 0.70 (excepto función emocional 0.64) - Al utilizar el coeficiente de correlación de Pearson para analizar las relaciones entre diferentes dimensiones de la escala EORTC QLQ-C15-PAL, se encontraron fuertes relaciones ($r > 0.07$) entre fatiga, dolor y pérdida de apetito. Entre las escalas de síntomas, la escala de fatiga se correlacionó actitudinalmente con la pérdida de apetito y el dolor, que mostraron coeficientes de correlación de $r = 0.632$ y $r = 0.635$, respectivamente.

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
(EORTC QLQ-C15-PAL)						
Ozcelik et al. (2016)	150 pacientes con cáncer avanzado. Diagnóstico primario; gastrointestinal (32.7%), osteosarcoma (10%), mama (19.3%), genitourinario (14%), pulmón (5.3%), piel (1.3%) y otros (17.3%).	Turquia: El estudio se realizó en Izmir con pacientes que habían recibido tratamiento en la unidad de cuidados paliativos de un hospital universitario.	La versión turca consta de 15 preguntas, incluidas dos escalas funcionales, dos escalas de síntomas de varios elementos y cinco escalas de síntomas de un solo elemento. Las opciones de respuesta a estos 14 ítems se calificaron en una escala tipo Likert de 4 puntos: 1 (nada), 2 (un poco), 3 (bastante), y 4 (mucho). Un ítem adicional, la escala global de salud / calidad de vida, utiliza una escala numérica de 7 puntos del 1 (muy deficiente) al 7 (excelente). Los puntajes más altos para la escala funcional y la escala global de salud / calidad de vida indican una mejor calidad de vida.	1-Consistencia interna 2-Validez transcultural 3-Prueba de hipótesis 4-Sensibilidad al cambio	1-Inadecuada 2-Muy buena 3-Muy buena 4-Muy buena	- α Cronbach = 0.93- 0.98 (0.90 para toda la escala). - La mayoría de las áreas del cuestionario tenían correlaciones de bajas a moderadas con las demás. Las correlaciones moderadas fueron entre fatiga y función física ($r = -0.41$; $p < 0.05$) y entre insomnio y función emocional ($r = 0.53$; $p < 0.05$). Por el contrario, se encontraron correlaciones débiles entre náuseas / vómitos y pérdida de apetito ($r = 0.31$; $p < 0.05$) y entre insomnio y dolor ($r = 0.22$; $p < 0.05$).

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
(EORTC QLQ-C15-PAL)						
Raman et al. (2016)	282 pacientes con metástasis óseas. Cáncer primario; mama (23%), pulmón (27.7%), próstata (24.5%), renal (5.7%), colorrectal (7.4%) y otros (11.7%).	Canadá: 23 centros canadienses.	La EORTC QLQ-C15-PAL contiene cuatro escalas de varios elementos y seis elementos individuales. Cada ítem en ambos instrumentos se califica de 1 (nada) a 4 (mucho) en severidad, excepto por la escala general de calidad de vida, que se califica de 1 (muy mala) a 7 (excelente). Una puntuación alta sugiere una mayor angustia en la escala / ítem de síntomas, mientras que una puntuación alta en la escala / ítem funcional muestra una mayor capacidad funcional. Cada escala / elemento se convierte en una puntuación que va de 0 a 100.	1-Consistencia interna 2-Prueba de hipótesis 3-Sensibilidad al cambio	1-Muy buena 2-Muy buena 3-Muy buena	- α Cronbach = 0.34 – 0.85 al inicio del estudio y 0.45 – 0.88 en el seguimiento. - No hubo un alto grado de correlación entre las escalas QLQ-BM22 y las escalas EORTC QLQ-C15-PAL. La interferencia funcional tuvo el mayor grado de correlación con otras escalas, con coeficiente de Pearson $r = \geq 0.40$; $p < 0.05$ para las escalas físicas, globales, de dolor y fatiga al inicio del estudio, a los 10 días y a los 42 días de seguimiento. En la escala de aspectos psicosociales, se observó correlación con la escala emocional al inicio, a los 10 días y a los 42 días de seguimiento.

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
(EORTC QLQ-C15-PAL)						
Golčić et al. (2018)	151 pacientes con cáncer avanzado. Tipos de cáncer; cáncer de pulmón (22.5%), colorrectal (12.6%), páncreas (8.6%) y otros cánceres (90.7%). Metástasis confirmadas; pulmones (31.1%), hígado (26.5%), esquelético (20.5%) y otros (72.2%).	Croacia: Los pacientes fueron reclutados del Hospicio Marija Kozulic'.	La versión croata de la EORTC QLQ-C15-PAL consta de 15 ítems que se refieren a 2 escalas funcionales (funcionamiento físico y emocional), 2 escalas de síntomas (dolor y fatiga), 5 ítems de un solo síntoma (disnea, insomnio, pérdida de apetito, náuseas y vómitos y estreñimiento) y la escala global QL.	1-Consistencia interna 2-Validez transcultural 3-Prueba de hipótesis 4-Sensibilidad al cambio	1-Muy buena 2-Adecuada 3-Muy buena 4-Muy buena	- α Cronbach = ≥ 0.70 (excepto para la escala de fatiga 0.65) - Las correlaciones superiores a $r = 0.40$ fueron entre la escala de dolor y el insomnio ($r = 0.417$; $p < 0.05$), la escala de fatiga y la pérdida de apetito ($r = 0.428$; $p < 0.05$) y el ítem de fatiga "¿Se ha sentido débil?" y pérdida de apetito ($r = 0.423$; $p < 0.05$).
Rojas-Concha et al. (2020)	48 pacientes con cáncer avanzado y 35 profesionales sanitarios. Los diagnósticos de cáncer más frecuentes fueron; próstata (14.6%), estómago (10.4%) y mieloma múltiple (10.4%).	Chile: Los pacientes fueron reclutados de los servicios de cuidados paliativos de cuatro hospitales públicos de Santiago.	La EORTC QLQ-C15-PAL es una abreviatura de la versión chileno-española de la EORTC QLQ-C30. Contiene cuatro escalas; funcionamiento físico, funcionamiento emocional, náuseas / vómitos y fatiga. La escala de calidad de vida global de 2 ítems se reduce a un ítem y se prescinde por completo de cinco escalas / ítems.	1-Validez de contenido 2-Validez transcultural	1-Inadecuada 2-Dudosa	

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Multidimensional Fatigue Inventory (MFI-20)						
Hagelin et al. (2007)	584 participantes que se clasificaron en cuatro grupos: pacientes con cáncer paliativo, pacientes con cáncer que reciben radioterapia, pacientes ambulatorios sin cáncer y un grupo de personal del hospital. Tipos de cáncer en pacientes ingresados para cuidados paliativos; gastrointestinal (31%), ginecología (18%), pulmón (14%), urológico (10%), mama (6%), sangre / linfa (9%), SNC (5%), cabeza / cuello (4%) y desconocidos (3%).	Suecia: Los datos se recopilaron en una clínica de Estocolmo; al final del tratamiento de radioterapia, en la admisión a la clínica, en una visita a la clínica para pacientes ambulatorios y durante un evento educativo.	La versión sueca del Inventario de fatiga multidimensional (MFI-20) comprende cinco subescalas; fatiga general, fatiga física, fatiga mental, actividad reducida y motivación reducida. Cada subescala incluye 4 ítems con escalas Likert de cinco puntos, que van desde "Sí, eso es cierto" a "No, eso no es cierto". Las puntuaciones en cada subescala van de 4 a 20. Las puntuaciones más altas indican una mayor fatiga.	1-Consistencia interna 2-Validez transcultural 3-Prueba de hipótesis 4-Sensibilidad al cambio	1-Muy buena 2-Muy buena 3-Muy buena 4-Muy buena	- α Cronbach = 0.67-0.94. - Las puntuaciones en las subescalas estaban altamente correlacionadas ($p \leq 0.001$). Las subescalas de fatiga general y fatiga física fueron las más correlacionadas entre sí en todos los grupos. La correlación entre las subescalas del MFI-20 y los puntajes en el CR-10 osciló entre $r = 0.37-0.74$ ($p < 0.01$ para todas las subescalas y CR-10), siendo la subescala de fatiga general la que tiene la mayor correlación con la escala CR-10.

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
(EORTC QLQ LMC21)						
Blazeby et al. (2009)	356 pacientes con diagnóstico histológico de adenocarcinoma de colon o recto y metástasis hepáticas (263 grupo de hepatectomía y 93 grupo de tratamiento paliativo).	Reino Unido, Francia y Alemania: Los pacientes fueron reclutados en cuatro centros.	EORTC QLQ-LMC21 contiene 21 elementos. Se divide en cuatro escalas de varios ítems (fatiga, problemas nutricionales, dolor y problemas emocionales) y 9 ítems individuales (pérdida de peso, problemas del gusto, sequedad de boca, dolor de boca, neuropatía periférica, ictericia, problemas sociales, problemas del habla y problemas sexuales). Todos los ítems se puntúan en una escala Likert de 1 a 4 puntos ("nada", "un poco", "bastante", "mucho"). Las puntuaciones más bajas indican una mejor calidad de vida.	1-Validez de contenido 2-Consistencia interna 3-Validez de criterio 4-Sensibilidad al cambio	1-Inadecuada 2-Muy buena 3-Muy buena 4-Muy buena	- α Cronbach \Rightarrow 0.70 (excepto "problemas nutricionales" 0.69) - Las correlaciones entre escalas en la EORTC QLQ-LMC21 y la EORTC QLQ-C30 fueron generalmente débiles, excepto entre escalas similares en ambos instrumentos (por ejemplo, la nueva escala de fatiga EORTC QLQ-LMC21 y las escalas de fatiga, función física, rol y social en la EORTC QLQ-C30; $r = \geq 0.60$; $p < 0.05$).

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
(EORTC QLQ LMC21)						
Paradowska et al. (2017)	160 pacientes con cáncer colorrectal y metástasis hepáticas (63 grupo de hepatectomía y 97 grupo de tratamiento paliativo).	Polonia: Los pacientes fueron reclutados en tres hospitales de Cracovia.	La versión polaca del EORTC QLQ-LMC21. Como anteriormente	1-Validez de contenido 2-Consistencia interna 3-Validez transcultural 4-Fiabilidad 5-Validez de criterio 6-Sensibilidad al cambio	1-Inadecuada 2-Muy buena 3-Muy buena 4-Dudosa 5-Muy buena 6-Muy buena	- α Cronbach = 0.72 – 0.90. - ICC = 0.64 – 0.88 - En general, hubo correlaciones débiles ($r = <0.60$; $p <0.05$) entre los dos cuestionarios que confirman que la EORTC QLQ-LMC21 aborda problemas de salud no evaluados en la EORTC QLQ-C30. Sin embargo, se observaron correlaciones más altas ($r \Rightarrow 0.60$; $p <0.05$) entre escalas similares en ambos instrumentos, es decir, escala de dolor del módulo y escala de dolor del núcleo, escala de problemas emocionales del módulo y escala de función emocional del núcleo, y escala de fatiga del módulo y función física, función de rol y escalas de fatiga del núcleo.

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Functional Assessment of Chronic Illness Therapy-Palliative Care (FACIT-PAL)						
Lyons et al. (2009)	256 personas recién diagnosticadas con cáncer avanzado. Localización de la enfermedad primaria; gastrointestinal (42%), pulmón (33%), genitourinario (14%) y mama (11%).	Estados Unidos: Personas que participan en un ensayo controlado aleatorio (ECA) de una intervención de cuidados paliativos para el cáncer avanzado.	La Evaluación Funcional de la Terapia de Enfermedades Crónicas-Cuidados Paliativos (FACIT-Pal) es una medida de 46 ítems. El FACIT-Pal contiene el FACT-G de 27 ítems que mide 4 dominios de la calidad de vida; bienestar físico, bienestar social / familiar, bienestar emocional y bienestar funcional. Además del FACT-G de 27 ítems, el FACIT-Pal agrega 19 ítems que miden los resultados relacionados con los cuidados paliativos y no son específicos del cáncer. La puntuación varía de 0 a 184. Las puntuaciones más altas indican una mejor calidad de vida.	1-Validez estructural 2-Consistencia interna 3-Prueba de hipótesis 4-Sensibilidad al cambio	1-Muy buena 2-Inadecuada 3-Muy buena 4-Muy buena	- El CFA se limitó a una solución de 5 factores; "bienestar físico", "bienestar funcional", "bienestar emocional", "bienestar social" y "claridad de pensamiento y decisiones". - α Cronbach = 0.93 - Tanto la intensidad de los síntomas (ESAS; $r = -0.73$, $P < 0.001$) como los síntomas depresivos (CES-D; $r = -0.75$, $P < 0.001$) se relacionaron significativamente con todas las subescalas y escalas compuestas del FACIT-Pal.

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
(FACIT-PAL)						
Siebert et al. (2014)	Se reclutaron muestras de 154, 154 y 153 participantes en Sudáfrica, Uganda y Kenia, respectivamente, lo que arroja una muestra total de 461. No especifica los tipos de cáncer presentes en el estudio.	Sudáfrica, Kenia y Uganda: Los datos de Sudáfrica se recopilaron en dos hospicios y comunidades de servicios en áreas metropolitanas. Los datos de Uganda se recopilaron en un hospicio en la ciudad capital, Kampala, que ofrece atención domiciliaria y de día, y los datos de Kenia se recopilaron en dos sitios; un hospicio en la ciudad capital de Nairobi y otro hospicio rural ubicado en la región del Monte Kenia de la Provincia Central.	El FACIT-Pal agrega 19 elementos que miden los resultados relacionados con los cuidados paliativos y no son específicos del cáncer, que se generaron en función de las preocupaciones informadas a través de entrevistas con personas con cáncer avanzado.	1-Validez estructural 2-Consistencia interna 3-Validez transcultural 4-Prueba de hipótesis 5-Sensibilidad al cambio	1-Adecuada 2-Inadecuada 3-Adecuada 4-Muy buena 5-Muy buena	- Se realizó un EFA y el número de factores se estableció en tres; factor 1 "sentido de propósito y significado en la vida", factor 2 "síntomas físicos" y factor 3 "integración social" - α Cronbach = 0.81. - Las correlaciones de ítem-total corregidas variaron de $r = 0.24-0.71$ con solo tres correlaciones por debajo de $r = 0.30$ ("recibo apoyo emocional de mi familia", "mi familia ha aceptado mi enfermedad" y "estoy satisfecho con la comunicación familiar sobre mi enfermedad").

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
(FACIT-PAL)						
Siegert et al. (2014b)	Como anteriormente	Como anteriormente	El FACIT-Pal contiene el FACT-G de 27 ítems que mide cuatro dominios de la calidad de vida; bienestar físico (7 ítems), bienestar social / familiar (7 ítems), bienestar emocional (6 ítems) y bienestar funcional (7 ítems).	1-Validez estructural 2-Consistencia interna 3-Validez transcultural 4-Prueba de hipótesis 5-Sensibilidad al cambio	1-Adecuada 2-Inadecuada 3-Adecuada 4-Muy buena 5-Muy buena	- Se realizó un EFA y se estableció el número de factores en cuatro según la versión original; factor 1 "síntomas físicos", factor 2 "bienestar funcional", factor 3 "amigos y familia" y factor 4 "bienestar emocional". - α Cronbach = 0.90. - Las correlaciones de ítem-total corregidas variaron de $r = 0.24$ - 0.71 con solo tres correlaciones por debajo de $r = 0.30$ ("recibo apoyo emocional de mi familia", "mi familia ha aceptado mi enfermedad" y "estoy satisfecho con la comunicación familiar sobre mi enfermedad").
Shinall et al. (2018)	227 pacientes. No especifica los tipos de cáncer presentes en el estudio.	Estados Unidos: La muestra se reclutó de una clínica de cuidados paliativos para pacientes ambulatorios en un gran centro médico académico urbano, Nashville.	El FACIT-Pal de 14 ítems es una versión más corta del cuestionario FACIT-Pal de 46 ítems. Las respuestas para cada pregunta están en una escala de 5 puntos que van desde "nada" a "mucho". Las puntuaciones más bajas indican una mejor calidad de vida.	1-Consistencia interna 2-Prueba de hipótesis 3-Sensibilidad al cambio	1-Inadecuada 2-Adecuada 3-Adecuada	- α Cronbach = 0.76 - 2 ítems tuvieron correlaciones bajas con el resto del instrumento ($r = 0.09$ y $r = 0.21$) y la eliminación de cualquiera de ellos mejoró el α de Cronbach (a 0.78 y 0.77), lo que indica que estos ítems restaron valor a la consistencia interna del instrumento.

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Hospice Quality of Life Scale (HQLS)						
Kim et al. (2009)	180 pacientes con cáncer avanzado. Los diagnósticos primarios incluyeron cáncer de pulmón (19.2%), estómago (17.4%) y colorrectal (15.7%).	Corea: La encuesta se realizó en 29 instituciones de todo el país e incluyó pacientes de diversos orígenes demográficos.	La Escala de calidad de vida de hospicio (HQLS) consta de 40 preguntas dentro de 13 dominios categorizados en otras seis subescalas: "Físico", "Psicológico", "Espiritual", "Social", "Familia / Economía" y "Global". Las opciones de respuesta para los ítems 1-21 y 23-39 van de 1 (nada) a 5 (mucho) y las de los ítems 22 y 40 van de 1 (nada) a 10 (mucho). Los ítems 4-21 se refieren a la pregunta, "¿Cuántas molestias causaron los siguientes síntomas durante la semana pasada?" Una puntuación más baja significa una mejor calidad de vida.	1-Validez de contenido 2-Validez estructural 3-Consistencia interna 4-Fiabilidad 5-Prueba de hipótesis 6-Sensibilidad al cambio	1-Inadecuada 2-Adecuada 3-Inadecuada 4-Inadecuada 5-Muy buena 6-Muy buena	- Se realizó un EFA para decidir dominios y extraer 11 factores. - α Cronbach = 0.41-0.93. - Los coeficientes de correlación test-retest variaron de $r = 0.29$ - 0.83 . - HQLS se correlacionó significativamente con el "funcionamiento físico" del EORTC QLQ-C30 y con los "síntomas físicos" del MQLS. La puntuación del dominio en "síntomas psicológicos" de la HQLS estuvo altamente correlacionada con las subescalas "funcionamiento emocional" de la EORTC QLQ-C30 ($r = 0.73$) y "emoción" de la MQLS ($r = 0.71$). Para los dominios "espiritualidad", "valores de vida" y "autoestima", se encontraron correlaciones significativas con la subescala relacionada con la "cognición" en el MQLS ($r = 0.22$, $r = 0.23$ y $r = 0.19$, respectivamente).

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Perform Questionnaire (PQ)						
Baró et al. (2011)	437 pacientes con diagnóstico de cáncer. Localización del cáncer: mama (33.6%), pulmón (14.9%), colon (11.4%), recto (7.8%), ovario (6.2%), próstata (5%) y otros (22%).	España: Departamentos de oncología y cuidados paliativos de 50 hospitales públicos.	El cuestionario de desempeño (PQ) utiliza la perspectiva del paciente para evaluar la percepción de fatiga en el cáncer. Consta de 12 ítems con respuestas en una escala ordinal de cinco puntos. Los 12 ítems se dividen en tres dimensiones: "Limitaciones físicas", "Actividades de la vida diaria" y "Creencias y actitudes". Se obtienen una puntuación global y una puntuación tridimensional. Las puntuaciones más bajas indican una peor percepción de la fatiga por parte del paciente.	1-Consistencia interna 2-Fiabilidad 3-Prueba de hipótesis 4-Sensibilidad al cambio	1-Muy buena 2-Inadecuada 3-Muy buena 4-Muy buena	- α Cronbach = 0.94 - ICC = 0.83 - La comparación de las puntuaciones de PQ con las otras medidas de salud (FACT-F, NHP e índice de Karnofsky) reveló correlaciones más fuertes entre las puntuaciones de Perform y el FACT-F ($r = 0.80$; $p < 0.05$; dimensiones $r = 0.68-0.75$; $p < 0.05$) que entre las puntuaciones de desempeño y las dimensiones del NHP (desempeño general con NHP físico $r = 0.68$; $p < 0.05$; desempeño general con NHP psicológico $r = 0.56$; $p < 0.05$; dimensiones de desempeño con NHP físico $r = 0.57-0.67$; $p < 0.05$; dimensiones de desempeño con NHP psicológico $r = 0.44-0.55$; $p < 0.05$) o entre puntajes de desempeño e índice de Karnofsky ($r = 0.35$; $p < 0.05$; dimensiones $r = 0.23-0.26$; $p < 0.05$).

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Symptom Assessment Scale (SAS)						
Aoun et al. (2011)	572 pacientes con cáncer. Tipos de cáncer; mama (7%), colon (16%), gástrico (8%), cabeza / cuello (5%), pulmón (21%), próstata (11%), otro (26%) y no cáncer (5%).	Australia: Los pacientes fueron reclutados de cinco servicios de cuidados paliativos en Perth (servicios de cuidados paliativos de consulta, unidades de cuidados paliativos para pacientes hospitalizados y servicio de atención domiciliaria).	La escala de evaluación de síntomas (SAS) comprende siete elementos: dolor, insomnio, náuseas, problemas intestinales, problemas del apetito, problemas respiratorios y fatiga. Los pacientes califican el grado de angustia asociado con el síntoma utilizando una escala de calificación verbal de tipo Likert de 0 a 10, donde cero significa que no se experimenta ningún síntoma de angustia, "1" es un síntoma mínimo de angustia y "10" es el peor síntoma posible de angustia. El período de tiempo para calificar los síntomas es en las últimas 24 horas.	1-Consistencia interna 2-Validez transcultural 3-Fiabilidad 4-Prueba de hipótesis 5-Sensibilidad al cambio	1-Inadecuada 2-Muy buena 3-Dudosa 4-Muy buena 5-Muy buena	- α Cronbach = 0.64 – 0.92 - ICC = 0.84 – 0.92 - Hubo diferencias significativas entre el diagnóstico de cáncer y las puntuaciones de los síntomas intestinales (F = 2.334, p = 0.018), las puntuaciones de los síntomas respiratorios (F = 7.540, p <0.001) y el dolor (F = 4.326, p <0.001).

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Three-Levels-of-Needs Questionnaire (3LNQ)						
Johnsen et al. (2011)	74 pacientes con cáncer avanzado. Tumor primario; cerebro, tracto gastrointestinal, ginecológico, cabeza / cuello, mama, pulmón, hematológico, genitourinario y otros.	Dinamarca: Los participantes fueron seleccionados al azar entre pacientes con cáncer en estadio 3 ó 4 de los departamentos de oncología, medicina, urología y cirugía de dos hospitales de Copenhague.	El Cuestionario de los tres niveles de necesidades (3LNQ), basado en búsquedas bibliográficas, mide 12 necesidades importantes con tres enfoques diferentes cuando se utiliza como complemento de la EORTC QLQ-C30: (1) intensidad del problema, (2) carga del problema y (3) sintió necesidad. Además, también mide las necesidades paliativas más frecuentes, definidas como la necesidad de alivio de los problemas físicos, psicológicos y sociales de los pacientes con cáncer avanzado.	1-Validez de contenido 2-Prueba de hipótesis 3-Sensibilidad al cambio	1-Inadecuada 2-Muy buena 3-Muy buena	- Para el acuerdo entre observadores, la mediana kappa = 0.91. Por acuerdo entre paciente y observador; para "los ítems de intensidad del problema" la mediana kappa = 0.73, para "los ítems de carga del problema" la mediana kappa = 0.63 y para "los ítems de necesidad sentida" la mediana kappa = 0.26.

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
Four-Item Fatigue Scale (FIFS)						
Davis et al. (2013)	65 pacientes hospitalizados o ambulatorios con cáncer avanzado. No especifica los tipos de cáncer presentes en el estudio.	Estados Unidos: Pacientes reclutados en el departamento de medicina paliativa en entornos ambulatorios y hospitalarios en Cleveland.	La Escala de fatiga de cuatro elementos (FIFS) incluye una serie de preguntas que se refieren a la gravedad clasificada numérica y categóricamente, y si fue mejor, peor o igual a lo largo del tiempo. Las preguntas son; (1) ¿Tiene fatiga? "Nunca", "A veces", "Generalmente", "Siempre". (2) En una escala de 0 a 10 (0 es ausencia de fatiga y 10 es fatiga severa), ¿cómo calificaría su fatiga ahora mismo / hoy? (3) ¿Cuál es el nivel de su fatiga en este momento / hoy: ninguno, leve, moderado, severo? "Ninguno", "Leve", "Moderado", "Severo". Y (4) ¿Es tu fatiga hoy mejor, peor, igual que ayer?	1-Prueba de hipótesis 2-Sensibilidad al cambio	1-Adecuada 2-Adecuada	- El FIFS y el BFI se correlacionaron con ciertos ítems de la ICD-10. La correlación de Spearman entre el FIFS y el BFI fue $r = 0.69$ el día 1 y $r = 0.73$ 1 semana después ($p < 0.05$)

Estudio (autor y año)	Población / Tipo de cáncer	Localización	Descripción del instrumento	Propiedades de medición	Puntuación COSMIN	Valores de medición
(EORTC QLQ-FA12)						
Weis et al. (2017)	946 pacientes con varios diagnósticos tumorales. Grupo A (curativo; n = 311), Grupo B (paliativo; n = 222), Grupo C (superviviente I <12 meses; n = 212), Grupo D (superviviente II > 36 meses; n = 199). Ubicación del tumor; mama, cabeza / cuello, pulmón, colorrectal, próstata, ginecológico, hematológico, testicular, pancreático y otros.	Inglaterra, Francia, Alemania, Austria, Polonia, Países Bajos, Suecia, España, Italia, Egipto y Taiwán: El estudio se llevó a cabo como un estudio multicéntrico internacional que incluyó 17 centros en 11 países europeos y no europeos.	El Cuestionario de 12 ítems de la Organización Europea para la Investigación y el Tratamiento del Cáncer de Calidad de Vida (EORTC QLQ-FA13) evalúa los aspectos físicos, cognitivos y emocionales de la fatiga relacionada con el cáncer. Este módulo se basa en un concepto multidimensional de fatiga que incluye 13 ítems (11 ítems asignados a un dominio físico, emocional y cognitivo y 2 ítems individuales sobre la interferencia de la fatiga relacionada con el cáncer con las actividades diarias y las secuelas sociales de la fatiga).	1-Validez de contenido 2-Validez estructural 3-Consistencia interna 4-Fiabilidad	1-Inadecuada 2-Adecuada 3-Inadecuada 4-Dudosa	- Se realizó un CFA. Los 3 factores identificados fueron; factor 1 "físico", factor 2 "emocional" y factor 3 "cognitivo". - α Cronbach => 0.70. - ICC: Rt1-t2 = 0.905– 0.921.

Abreviaturas: (HQLI) Hospice Quality of Life Index; (VAS) visual analogue scale; (EORTC QLQ-C30) European Organization for Research and Treatment of Cancer Core Quality of Life Questionnaire 30 items; (Global QoL) Global Quality of Life; (ICC) Intraclass Correlation Coefficient; (QL) Global Quality of Life; (QoL) Quality of Life; (HADS) Hospital Anxiety and Depression Scale; (BPI) Brief Pain Inventory; (BPI-I) Brief Pain Inventory- Intensity; (BPI-F) Brief Pain Inventory- Function; (LC13) Lung Cancer specific questionnaire 13 items; (MQLS) McMaster Quality of Life Scale; (K-MQLS) Korean- McMaster Quality of Life Scale; (QLI) Spitzer Quality of Life Index; (ESAS) Edmonton Symptom Assessment System; (RSCL) Rotterdam Symptom Check List; (SDS) symptom distress score; (ESAS-Thai) Thai Version of the Edmonton Symptom Assessment Scale; (M-ESAS) Modified version of the Edmonton Symptom Assessment Scale; (ESAS-r) Edmonton Symptom Assessment System-revised; (PPS) Palliative Performance Scale; (K-ESAS) Korean version of the Edmonton Symptom Assessment Scale; (K-MDASI) Korean versions of the M. D. Anderson Symptom Inventory; (K-HADS) Korean-Hospital Anxiety and Depression Scale; (ESAS-CS) Constipation and Sleep-Edmonton Symptom Assessment System; (ESAS-r-CS) The revised version of the Edmonton Constipation and Sleep Symptom Assessment System; (MSAS) Memorial Symptom Assessment Scale; (ESS) Epworth Sleepiness Scale; (MDASI) M. D. Anderson Symptom Inventory; (F-ESAS) French version of the Edmonton Symptom Assessment System; (ESAS: AM) Edmonton Symptom Assessment System: Ascites Modification; (EORTC QLQ-PAN26) European Organization for Research and Treatment of Cancer Quality of Life Questionnaire, 26-items pancreatic cancer module; (ESAS:AM-24h) Edmonton Symptom Assessment System: Ascites Modification - 24 hours; (PSI) Person Separation Index; (ESAS12-F) French Version of the Modified Edmonton Symptom Assessment Scale; (CFS) Cancer Fatigue Scale; (MMS) Mini Mental State; (BFI-J) The Japan version of the Brief Fatigue Inventory; (POMS) Profile of Mood States; (BFI-Gr) The Greek version of the Brief Fatigue Inventory; (FIFS) Four-Item Fatigue Scale; (ICD-10) International Classification of Diseases-Tenth Revision; (CAMPAS-R) Cambridge Palliative Assessment Schedule; (PQLI) Palliative Care Quality of Life Instrument; (AQEL) Palliative Care Quality of Life Assessment; (PNPC) Palliative Care for Cancer Patients; (HRQL) health-related quality of life; (ADL) activities of daily living; (IADL) instrumental activities of daily living; (COOP) questionnaire of sheets COOP/WONCA Charts; (PNPC-sv) Palliative Care of Cancer Patients- short version; (EORTC QLQC15-PAL) European Organization for Research and Treatment of Cancer Quality of Life 15 items Questionnaire for Palliative Care; (INCan) National Cancer Institute; (UCP) palliative care unit; (KPS) Karnofsky performance status; (BPI-SF) Brief Pain Inventory - Short Form; (BDI) Beck Depression Inventory; (GFI) goodness-of-fit index; (CFI) comparative fit index; (NFI) normed fit index; (NNFI) non-normed fit index; (ECOG) Eastern Cooperative Oncology Group; (QLQ-BM22) European Organization for Research and Treatment of Cancer Quality of Life Questionnaire - Bone Metastases 22-items; (MFI-20) Multidimensional Fatigue Inventory- 20 items; (CR-10) Borg Category Ratio scale; (EORTC QLQ-LMC21) European Organization for Research and Treatment of Cancer Colorectal Liver Metastases- 21items; (FACIT-Pal) Functional Assessment of Chronic Illness Therapy- Palliative Care; (FACT-G) Functional Assessment of Cancer Therapy- General; (CES-D) Center for Epidemiological Studies-Depression; (HQLS) Hospice Quality of Life Scale; (PQ) Perform Questionnaire; (FACT-F) Functional Assessment of Cancer Therapy-Fatigue; (NHP) Nottingham Health Profile; (SAS) Symptom Assessment Scale; (3LNQ) Three Levels of Needs Questionnaire; (EORTC QLQ-FA12) European Organization for Research and Treatment of Cancer Quality of Life 12 items Questionnaire.

3.1 Propiedades psicométricas

Las principales propiedades evaluadas fueron la consistencia interna, la validez transcultural, la prueba de hipótesis y la sensibilidad al cambio. Los resultados de la lista de verificación de cuatro puntos COSMIN de instrumentos para medir la fatiga en pacientes tratados con CP se muestran en la tabla 14.

En la tabla 15 se muestran los datos de los instrumentos para medir la FRC en CP utilizando la lista de verificación COSMIN y el método GRADE.

Instrumento	Artículo	Validez de contenido		Validez estructural		Consistencia interna		Validez transcultural / invariancia de medición		Fiabilidad		Validez de criterio		Prueba de hipótesis		Sensibilidad al cambio	
		M	Q	M	Q	M	Q	M	Q	M	Q	M	Q	M	Q	M	Q
McMaster Quality of Life Scale (MQLS)	Sterkenburg et al. (1996)					D	-			I	-			V	+	V	+
	Kim et al. (2006)	I	+ -			V	-	I	-	I	-			A	+	A	+
Edmonton Symptom Assessment System (ESAS)	Philip et al. (1998)							I	?					A	?	A	?
	Carvajal et al. (2011)	I	+ -	A	?	V	+	A	+	V	-			V	?	V	?
	Chinda et al. (2011)	I	-			I	-	I	+								
	Claessens et al. (2011)	I	-					I	+					A	+	A	+
	Carvajal et al. (2013)					V	+	I	+	V	+			V	?	V	?
	Kwon et al. (2013)	I	-			I	+	A	+	V	+			V	+	V	+
	Hannon et al. (2015)					V	+			D	-			V	?	V	?
	Hannon et al. (2015b)					V	+							V	?	V	?
	Paiva et al. (2015)	I	-			V	+	V	-	D	+			V	+	V	+
	Yokomichi et al. (2015)	I	-			I	+	V	+	D	+			V	?	V	?
Gretarsdottir et al. (2016)					I	+							V	?	V	?	

Instrumento	Artículo	Validez de contenido		Validez estructural		Consistencia interna		Validez transcultural / invariancia de medición		Fiabilidad		Validez de criterio		Prueba de hipótesis		Sensibilidad al cambio	
		M	Q	M	Q	M	Q	M	Q	M	Q	M	Q	M	Q	M	Q
Edmonton Symptom Assessment System (ESAS)	Nekolaichuk et al. (2017)	I	-					I	?								
	Pautex et al. (2017)	I	-			V	+							A	?	A	?
	Mori et al. (2018)	I	-	A	?	I	+	V	+	D	+			V	?	V	?
	Mori et al. (2018b)	I	-			I	+	V	+					V	?	V	?
	Sprague et al. (2018)			V	+									V	?	V	?
	Tricou et al. (2019)							I	?								
Cancer Fatigue Scale (CFS)	Okuyama et al. (2000)			I	?	V	+	V	+	D	-			V	?	V	?
Brief Fatigue Index (BFI)	Okuyama et al. (2003)			D	?	V	+	V	+					V	+	V	+
	Mystakidou et al. (2008)			A	?	V	+	D	+	D	+			V	?	V	?
	Davis et al. (2013)													A	?	A	?
Cambridge Palliative Assessment Schedule (CAMPAS-R)	Ewing et al. (2004)					V	+							V	?	V	?
Palliative Care Quality of Life Instrument (PQLI)	Mystakidou et al. (2004)	I	+ -	A	?	V	+	D	+					V	?	V	?

Instrumento	Artículo	Validez de contenido		Validez estructural		Consistencia interna		Validez transcultural / invariancia de medición		Fiabilidad		Validez de criterio		Prueba de hipótesis		Sensibilidad al cambio	
		M	Q	M	Q	M	Q	M	Q	M	Q	M	Q	M	Q	M	Q
Problems and Needs in Palliative Care Questionnaire (PNPC)	Osse et al. (2004)	I	+			D	-	I	+					A	+	A	+
	Osse et al. (2007)		-			I	-					I	+			I	?
	Wang et al. (2019)	I	+	A	+	I	-	A	-					V	+	V	+
European Organization for Research and Treatment of Cancer Quality of Life 15 items Questionnaire for Palliative Care (EORTC QLQ-C15-PAL)	Groenvold et al. (2006)	I	+							I	-	I	+			I	?
	Suarez Del Real et al. (2010)					I	-	D	+	I	NS			V	?	V	?
	Miyazaki et al. (2011)					I	-	D	+	I	+	I	+			I	?
	Shin et al. (2011)					V	-	V	-					V	?	V	?
	Leppert et al. (2012)					D	+	V	+					V	?	V	?
	Arraras et al. (2013)					V	-	V	-					V	?	V	?
	Nunes et al. (2014)			V	+	V	-	V	+			I	-			I	+
	Alawneh et al. (2015)			A	+	I	+	A	-					V	?	V	?
	Miyashita et al. (2015)			A	?	I	+	V	+	I	-			V	?	V	?
	Zhang et al. (2015)					I	-	V	-			I	+			I	?
Ozcelik et al. (2016)					I	+	V	-					V	?	V	?	

Instrumento	Artículo	Validez de contenido		Validez estructural		Consistencia interna		Validez transcultural / invariancia de medición		Fiabilidad		Validez de criterio		Prueba de hipótesis		Sensibilidad al cambio	
		M	Q	M	Q	M	Q	M	Q	M	Q	M	Q	M	Q	M	Q
European Organization for Research and Treatment of Cancer Quality of Life 15 items Questionnaire for Palliative Care (EORTC QLQ-C15-PAL)	Raman et al. (2016)					V	-							V	+	V	+
	Golčić et al. (2018)					V	-	A	+					V	?	V	?
	Rojas-Concha et al. (2020)	I	+							D	?						
Multidimensional Fatigue Inventory (MFI-20)	Hagelin et al. (2007)					V	-	V	-					V	?	V	?
European Organization for Research and Treatment of Cancer Colorectal Liver Metastases (EORTC QLQ-LMC21)	Blazeby et al. (2009)	I	+			V	-					V	-			V	+
	Paradowska et al. (2017)	I	-			V	+	V	-	D	-	V	-			V	?
Functional Assessment of Chronic Illness Therapy-Palliative Care (FACIT-PAL)	Lyons et al. (2009)			V	?	I	+							V	+	V	+
	Siegert et al. (2014)			A	?	I	+	A	+					V	?	V	?
	Siegert et al. (2014b)			A	?	I	+	A	+					V	?	V	?
	Shinall et al. (2018)					I	+							A	+	A	+

Instrumento	Artículo	Validez de contenido		Validez estructural		Consistencia interna		Validez transcultural / invariancia de medición		Fiabilidad		Validez de criterio		Prueba de hipótesis		Sensibilidad al cambio	
		M	Q	M	Q	M	Q	M	Q	M	Q	M	Q	M	Q	M	Q
Hospice Quality of Life Scale (HQLS)	Kim et al. (2009)	I	+ -	A	?	I	-			I	-			V	?	V	?
Perform Questionnaire (PQ)	Baró et al. (2011)					V	+			I	+			V	?	V	?
Symptom Assessment Scale (SAS)	Aoun et al. (2011)					I	-	V	-	D	+			V	+	V	+
Three-Levels-of-Needs Questionnaire (3LNQ)	Johnsen et al. (2011)	I	+ -											V	?	V	?
Four-Item Fatigue Scale (FIFS)	Davis et al. (2013)													A	?	A	?
European Organization for Research and Treatment of Cancer Quality of Life 12 items Questionnaire (EORTC QLQ-FA12)	Weis et al. (2017)	I	+ -	A	+	I	+			D	+						

El error de medición se omitió en la tabla porque no se evaluó en ningún estudio.

M: Calidad metodológica del estudio calificada como I = Inadecuado, D = Dudoso, A = Adecuado, V = Muy bueno, casillas vacías = no informado

Q: Calidad de los resultados calificados como + = calificación suficiente, ? = calificación indeterminada, ± = calificación inconsistente, - = calificación insuficiente, casillas vacías = no reportado

NS: No especificado

Tabla 15. Resumen de hallazgos: calidad metodológica, calidad de los resultados y calidad de la evidencia reportada por propiedad e instrumento de medición.

Instrumento	Validez de contenido			Validez estructural			Consistencia interna			Validez transcultural / invariancia de medición			Fiabilidad			Validez de criterio			Prueba de hipótesis			Sensibilidad al cambio		
	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE
Hospice Quality of Life Index (HQLI)	I	+	VL																D	?	M	D	?	M
European Organization for Research and Treatment of Cancer Core Quality of Life Questionnaire (EORTC QLQ-C30)				D	?	M	V	-	H	D	+	L	I	-	L				A	-	L	A	-	L
McMaster Quality of Life Scale (MQLS)	I	+	VL				D	-	M	I	-	VL	I	-	L				A	+	M	A	+	M

Instrumento	Validez de contenido			Validez estructural			Consistencia interna			Validez transcultural / invariancia de medición			Fiabilidad			Validez de criterio			Prueba de hipótesis			Sensibilidad al cambio		
	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE
Edmonton Symptom Assessment System (ESAS)	I	-	L	A	?	M	I	-	M	I	-	L	D	-	M				A	?	H	A	?	H
Cancer Fatigue Scale (CFS)				I	?	VL	V	+	H	V	+	H	D	-	L				V	?	H	V	?	H
Brief Fatigue Index (BFI)				D	?	L	V	+	H	D	+	M	D	+	L				A	?	M	A	?	M
Cambridge Palliative Assessment Schedule (CAMPAS-R)							V	+	H										V	?	H	V	?	H
Palliative Care Quality of Life Instrument (PQLI)	I	+	VL	A	?	M	V	+	H	D	+	M							V	?	H	V	?	H

Instrumento	Validez de contenido			Validez estructural			Consistencia interna			Validez transcultural / invariancia de medición			Fiabilidad			Validez de criterio			Prueba de hipótesis			Sensibilidad al cambio		
	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE
Problems and Needs in Palliative Care Questionnaire (PNPC)	I	+	L	A	+	M	I	-	L	I	-	VL				I	+	VL	A	+	H	I	?	M
European Organization for Research and Treatment of Cancer Quality of Life 15 items Questionnaire for Palliative Care (EORTC QLQ-C15-PAL)	I	+	L	A	?	H	I	-	L	D	-	M	I	-	L	I	-	L	V	?	H	I	?	L
Multidimensional Fatigue Inventory (MFI-20)							V	-	H	V	-	H				V	?	H	V	?	H			

Instrumento	Validez de contenido			Validez estructural			Consistencia interna			Validez transcultural / invariancia de medición			Fiabilidad			Validez de criterio			Prueba de hipótesis			Sensibilidad al cambio		
	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE
European Organization for Research and Treatment of Cancer Colorectal Liver Metastases (EORTC QLQ-LMC21)	I	-	L				V	-	M	V	-	H	D	-	L	V	-	H				V	?	H
Functional Assessment of Chronic Illness Therapy-Palliative Care (FACIT-PAL)				A	?	H	I	+	L	A	+	H							A	?	M	A	?	M
Hospice Quality of Life Scale (HQLS)	I	+	VL	A	?	M	I	-	VL				I	-	VL				V	?	H	V	?	H
Perform Questionnaire (PQ)							V	+	H				I	+	VL				V	?	H	V	?	H

Instrumento	Validez de contenido			Validez estructural			Consistencia interna			Validez transcultural / invariancia de medición			Fiabilidad			Validez de criterio			Prueba de hipótesis			Sensibilidad al cambio		
	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE	M	Q	QE
Symptom Assessment Scale (SAS)							I	-	VL	V	-	H	D	+	M				V	+	H	V	+	H
Three-Levels-of-Needs Questionnaire (3LNQ)	I	+	VL																V	?	H	V	?	H
Four-Item Fatigue Scale (FIFS)																			A	?	M	A	?	M
European Organization for Research and Treatment of Cancer Quality of Life 12 items Questionnaire (EORTC QLQ-FA12)	I	+	VL	A	+	M	I	+	VL				D	+	L									

El error de medición se omitió en la tabla porque no se evaluó en ningún estudio.

M: Calidad metodológica del estudio calificada como I = Inadecuado, D = Dudoso, A = Adecuado, V = Muy bueno, casillas vacías = no informado

Q: Calidad de los resultados calificados como + = calificación suficiente, ? = calificación indeterminada, ± = calificación inconsistente, - = calificación insuficiente, casillas vacías = no reportado

QE: Calidad de la evidencia calificada como H = Alta, M = Moderada, L = Baja, VL = Muy baja

3.1.1 Hospice Quality of Life Index (HQLI)

La validez de contenido, la prueba de hipótesis y la capacidad de respuesta se evaluaron en un solo estudio (24). Para la validez de contenido, la calidad metodológica se clasificó como "inadecuada". Para la prueba de hipótesis, la calificación fue "dudosa" y las correlaciones de Pearson fueron más altas para los ítems relacionados con "contacto físico con otros", "qué tan bien come el paciente" y "qué tan bien duerme el paciente" ($r = 0.47$, $r = 0.44$ y $r = 0.42$, respectivamente).

3.1.2 European Organization for Research and Treatment of Cancer Core Quality of Life Questionnaire (EORTC QLQ C30)

Este instrumento se incluyó en 5 estudios (25-29) y se evaluaron muchas de las propiedades psicométricas, excepto la validez de contenido, el error de medición y la validez de criterio. La validez estructural obtuvo una calificación de "dudosa" o "adecuada" para los 2 estudios donde se evaluó (27,28), obteniendo una estructura de 6 factores (27) y otra de 2 (28). Para la consistencia interna, la calificación fue "muy buena", y el resultado fue un alfa de Cronbach > 0.70 para todos los estudios donde se evaluó esta propiedad (25-27,29), pero con algunas excepciones, ya que se obtuvo un alfa de Cronbach de < 0.70 en las escalas de funcionamiento (rol y cognitivo) (25); en las subescalas de funcionamiento emocional y función cognitiva (26); en las escalas de funcionamiento físico y síntomas de fatiga; los síntomas náuseas/vómitos y dolor (27), y en la escala de funcionamiento cognitivo (29). La fiabilidad test-retest se midió para un solo estudio y se clasificó como "inadecuada", resultando un ICC > 0.70 (excepto para las subescalas de síntomas: náuseas/vómitos, donde ICC = 0.63) (26). La prueba de hipótesis fue "adecuada" o "muy buena" en todos los estudios, con correlaciones significativas entre diferentes escalas (tabla 13) (25-29).

3.1.3 McMaster Quality of Life Scale (MQLS)

MQLS se evaluó en 2 de los estudios (30,31) para la mayoría de las propiedades psicométricas, excepto la validez estructural, el error de medición y la validez de criterio. La calificación de la calidad metodológica de este

instrumento para el único estudio donde se evaluó (31) fue "inadecuada". Los resultados con respecto a la consistencia interna para los 2 estudios donde estuvo presente fueron "dudosos" o "muy buenos", resultando valores para alfa de Cronbach < 0.70 . La calidad metodológica para la fiabilidad fue "inadecuada" para los 2 estudios, obteniendo un ICC < 0.70 en ambos casos. La prueba de hipótesis fue "adecuada" o "muy buena" (30,31), revelando correlaciones moderadas con Spitzer Quality of Life Index (QLI) y Korean- McMaster Quality of Life Scale (K-MQLS).

3.1.4 *Edmonton Symptom Assessment System (ESAS)*

Este instrumento fue el de mayor representatividad en cuanto a número de estudios, 15 en total (32-46). Se midieron todas las propiedades psicométricas, excepto el error de medición y la validez de criterio. La validez de contenido fue catalogada como "inadecuada" para todos los estudios donde fue evaluada. La validez estructural se calificó como "adecuada" o "muy buena" en 3 de los estudios en los que estuvo presente (33,44,45).

Para el primer estudio, se llevó a cabo un análisis factorial exploratorio (EFA) que encontró 3 dominios centrales: "dominio físico suave" (el primer dominio, descrito como "síntomas físicos suaves" incluye náuseas, apetito, somnolencia y debilidad), "dominio emocional" (el segundo dominio, descrito como "componentes emocionales" incluye ansiedad, depresión u otros síntomas como sensación de bienestar o dificultad para dormir) y "dominio físico duro" (el tercer dominio, descrito como "síntomas físicos duros" incluye dolor y dificultad para respirar) (33).

Para el segundo estudio, el mismo autor utiliza dos versiones de ESAS: una que califica los síntomas centrándose únicamente en las 24 horas previas (ESAS:AM- 24h) y que no mide validez estructural, y otra que examina los síntomas en el momento actual (ESAS:AM), en la cual el autor sí realizó un EFA que identificó 2 factores: "síntomas físicos" y "síntomas psicológicos" (44).

Finalmente, el tercero, empleó un modelo Rasch de crédito parcial. En el modelo de Rasch inicial se mostró un índice de separación de personas (PSI) de 0.77, lo que indicaba una fiabilidad satisfactoria, pero el ajuste general del modelo fue pobre ($X^2 = 155.41$; $p < 0.01$), y todos los elementos mostraron umbrales

desordenados (45). La consistencia interna fue calificada como "inadecuada" o "muy buena" para todos los estudios donde estuvo presente (33,34,36–41,43,44) y obtuvo valores de alfa de Cronbach > 0.70 en todos ellos. La propiedad psicométrica fiabilidad fue calificada como "dudosa" o "muy buena" para todos los estudios donde fue evaluada. La prueba de hipótesis se calificó como "adecuada" o "muy buena" (32,33,35–41,43–45). Los resultados de esta propiedad psicométrica para cada estudio se muestran en la tabla 13.

3.1.5 *Cancer Fatigue Scale (CFS)*

Este instrumento solo fue evaluado en un estudio (47). La validez estructural resultó "inadecuada", se realizó un EFA y se identificaron 3 factores mediante la prueba de Scree: "subescala física" (la primera subescala constaba de 7 ítems que se relacionaban con las sensaciones físicas de fatiga: cansarse fácilmente, tener ganas de acostarse, exhausto, pesado y cansado, harto, reacio y fatigado no sabes que hacer contigo mismo); "subescala afectiva" (la segunda subescala constaba de 4 ítems que evaluaban actividad afectiva: sentimiento energético, interés en algo, capacidad para concentrarse y animarse a hacer algo), y "subescala cognitiva" (la tercera subescala constaba de 4 ítems que evaluaban la atención y la memoria: descuido, errores al hablar, olvidadizo y el pensamiento se ha vuelto más lento).

La consistencia interna fue calificada como "muy buena" y obtuvo valores para alfa de Cronbach de 0.88. La fiabilidad fue "dudosa" y el promedio para los coeficientes de correlación test-retest fue 0.69 ($p < 0.01$). Para la prueba de hipótesis la calidad metodológica resultó "muy buena", la validez convergente de CFS fue demostrada por las correlaciones de Pearson con los puntajes de la visual analogue scale (VAS). En la asociación entre la escala de fatiga por cáncer, el estado funcional, Hospital Anxiety and Depression Scale (HADS) y Mini Mental State (MMS), la puntuación del estado funcional se asoció significativamente con las puntuaciones de la subescala física y afectiva.

3.1.6 *Brief Fatigue Index (BFI)*

BFI se incluyó en 3 estudios (48–50). Se midieron prácticamente todas las propiedades psicométricas, excepto la validez de contenido, el error de medición y la validez de criterio. La validez estructural en relación con su calidad metodológica se calificó como “dudosa” o “adecuada” con una estructura de solución unifactorial (48,49). La consistencia interna fue “muy buena” y se obtuvieron valores para alfa de Cronbach de 0.954 (49) y 0.96 (48), respectivamente. La fiabilidad test-retest se calificó como “dudosa” y solo se evaluó en un estudio (49), resultando un valor de 0.901. La prueba de hipótesis se calificó como “adecuada” o “muy buena” para los 3 estudios cuando se evaluó BFI.

3.1.7 *Cambridge Palliative Assessment Schedule (CAMPAS-R)*

CAMPAS-R se evaluó para un único estudio (51). La consistencia interna fue “muy buena” y obtuvo valores alfa de Cronbach de 0.77 (para severidad) y 0.80 (para interferencia), respectivamente. La prueba de hipótesis también fue calificada como “muy buena”, encontrando correlaciones muy fuertes entre los ítems de disnea en el EORTC QLQ-C30 y el CAMPAS-R y entre los ítems de estreñimiento en las dos medidas, así como entre las puntuaciones de ansiedad del paciente en el CAMPAS -R y la subescala de ansiedad HADS.

3.1.8 *Palliative Care Quality of Life Instrument (PQLI)*

PQLI estuvo presente en solo un estudio (52). Para la validez de contenido, la calidad metodológica se calificó como “inadecuada”. La validez estructural resultante fue “adecuada”, realizándose un EFA que incluyó 7 factores: “actividad”, “autocuidado”, “estado de salud”, “elección del tratamiento”, “apoyo”, “comunicación” y “afecto psicológico”. La consistencia interna fue catalogada como “muy buena”, resultando valores de alfa de Cronbach ≥ 0.70 . Finalmente, la prueba de hipótesis también resultó como “muy buena”, y las correlaciones más fuertes se obtuvieron entre PQLI y Palliative Care Quality of Life Assessment (AQEL) ($r = 0.44-0.94$; $p < 0.05$).

3.1.9 Problems and Needs in Palliative Care Questionnaire (PNPC)

En 3 de los estudios (53-55) se evaluaron prácticamente todas las propiedades psicométricas de este instrumento, excepto el error de medición y la fiabilidad. La validez de contenido se midió y calificó como "inadecuada" en 2 de los 3 estudios (54,55). La validez estructural fue calificada en solo uno de estos como "adecuada" (55), realizándose un análisis factorial confirmatorio (CFA), obteniéndose 2 factores. La consistencia interna se clasificó como "inadecuada" o "dudosa" y se evaluó en todos los artículos, resultando valores alfa de Cronbach < 0.70 para algunas de las subescalas en los 3 estudios (53-55). Para este instrumento, la validez de criterio se evaluó en un único artículo que comparó una versión corta con la versión original, obteniendo una calificación de "inadecuada" (54). En cuanto a la prueba de hipótesis, fue calificada como "adecuada" o "muy buena" para los 2 estudios restantes, respectivamente (53,55). Los datos sobre las correlaciones en ambos estudios se muestran en la tabla 13.

3.1.10 European Organization for Research and Treatment of Cancer Quality of Life 15-item Questionnaire for Palliative Care (EORTC QLQ-C15-PAL)

EORTC QLQ-C15-PAL fue el segundo instrumento incluido en el mayor número de estudios, 14 en total (56-69), y en el que se midieron más propiedades psicométricas: validez de contenido, validez estructural, consistencia interna, validez transcultural, fiabilidad, validez de criterio, prueba de hipótesis y sensibilidad al cambio (excepto el error de medición).

La validez de contenido fue catalogada como "inadecuada" para los 2 estudios donde fue evaluada (56,69). La validez estructural se calificó como "adecuada" o "muy buena" y estuvo presente en 3 de los estudios (62-64). Para los 2 primeros estudios se realizó un CFA (62,63), resultando una solución de 3 factores: "estado de salud global", "escala funcional" y "escala de síntomas" (62); y 5 factores: "emocional", "fatiga", "dolor", "físico" y "síntomas" (63). Para el último de los estudios (64), se ejecutó un EFA y se estableció en 9 el número de factores según la versión original.

La consistencia interna se calificó como "inadecuada", "dudosa" o "muy buena" para los estudios donde se evaluó (57-68), resultando valores alfa de

Cronbach < 0.70 para 8 de los estudios (57–59, 61, 62, 65, 67, 68) y valores alfa de Cronbach > 0.70 para los 4 restantes (60,63,64,66). La fiabilidad fue calificada como “inadecuada” para los 4 estudios donde estuvo presente (56–58,64).

Para este instrumento se midió la validez de criterio, la cual fue calificada como “inadecuada”. EORTC QLQ-C15-PAL (versión reducida del original) se comparó con su versión original, EORTC QLQ-C30, en 4 de los estudios (56,58,62,65). Las correlaciones entre las puntuaciones predichas y observadas oscilaron entre 0.87 ($p < 0.05$) (escala de función cognitiva) y 0.97 ($p < 0.05$) (escala de fatiga) (56). La correlación ítem-escala para los ítems de EORTC QLQ-C15-PAL varió entre 0.74 y 0.95 (58), se obtuvieron correlaciones positivas y de moderadas a perfectas entre EORTC QLQ-C15-PAL y EORTC QLQ-C30 en un rango de 0.30-1 (62) y la proporción de varianza (R^2) osciló entre 0.848 y 0.903, lo que mostró que las puntuaciones de la subescala EORTC QLQ-C15-PAL explicaban entre el 84.8 % y el 90.3 % de la distribución de puntuaciones de EORTC QLQ-C30 original (65).

La prueba de hipótesis fue calificada como “muy buena” para el total de artículos donde se evaluó QLQ-C15-PAL (57,59–61,63,64,66–68). Los datos sobre la prueba de hipótesis se presentan en la tabla 13.

3.1.11 Multidimensional Fatigue Inventory (MFI-20)

Las propiedades psicométricas de MFI-20 se evaluaron en un solo artículo (70). Dentro de las propiedades de medición evaluadas, la consistencia interna fue calificada como “muy buena” y obtuvo valores de alfa de Cronbach entre 0.67 y 0.94. La prueba de hipótesis fue calificada como “muy buena”, siendo las subescalas de fatiga general y fatiga física las más correlacionadas entre sí en todos los grupos.

3.1.12 European Organization for Research and Treatment of Cancer Colorectal Liver Metastases (EORTC QLQ-LMC21)

EORTC QLQ-LMC21 se evaluó en 2 estudios (71,72), en los que se midieron un número considerable de propiedades psicométricas (validez de contenido, consistencia interna, validez transcultural, fiabilidad, validez de criterio y

sensibilidad al cambio). La validez de contenido se calificó como "inadecuada" en ambos estudios. La consistencia interna también fue evaluada en ambos artículos y fue calificada desde el punto de vista metodológico como "muy buena", con valores alfa de Cronbach < 0.70 (71) y entre 0.72 y 0.90, respectivamente (72). La fiabilidad test-retest se calificó como "dudosa" y se midió solo en uno de los estudios (72) obteniendo un ICC entre 0.64 y 0.88. La validez de criterio se evaluó comparando el instrumento EORTC QLQ-LMC21 (versión reducida del original) con EORTC QLQ-C30 (versión original) para ambos estudios (71,72), calificando su calidad metodológica como "muy buena" y resultando correlaciones significativamente débiles entre ambas herramientas.

3.1.13 Functional Assessment of Chronic Illness Therapy-Palliative Care (FACIT-PAL)

FACIT-PAL estuvo presente en 3 de los estudios (73–75). Dentro de las propiedades psicométricas evaluadas se midió la validez estructural para 2 de estos estudios (73,74), resultando una calificación de "adecuada" o "muy buena", y dando como resultado un CFA con una solución de 5 factores: "bienestar físico" (7 ítems: energía, náuseas, dificultad para satisfacer las necesidades familiares, dolor, efectos secundarios, sentirse enfermo y pasar tiempo en cama), "bienestar funcional" (7 ítems: capaz de trabajar, el trabajo es gratificante, disfrutar la vida, aceptar la enfermedad, dormir bien, disfrutar las cosas por diversión y calidad de vida), "bienestar emocional" (6 ítems: tristeza, afrontamiento, pérdida de esperanza, nervios, preocupación por morir y preocupación por empeoramiento), "bienestar social" (7 ítems: amigos cercanos, apoyo emocional familiar, apoyo de amigos, enfermedad aceptada por la familia, comunicación familiar, apoyo de pareja y vida sexual) y "claridad de pensamiento y decisiones" para el primer estudio (73); y una estructura de 3 y 4 factores para los dos subcuestionarios (FACIT-PAL 19 ítems y FACT-G 27 ítems) presentes en el estudio de Siegert et al., 2014 (74).

Se midió la consistencia interna para todos los estudios (73–75) y se catalogó como "inadecuada", resultando valores alfa de Cronbach entre 0.76 (75) y 0.93 (73). Para la prueba de hipótesis, la calificación fue "adecuada" o "muy buena"

para los 3 estudios. Los datos sobre las correlaciones encontradas en los artículos de investigación se presentan en la tabla 13.

3.1.14 Hospice Quality of Life Scale (HQLS)

La validez estructural de HQLS se catalogó como "adecuada" en uno de los estudios, en los que se realizó un EFA, resultando 11 factores: "síntomas psicológicos" (9 ítems: preocupación, ira, irritabilidad, ansiedad/inquietud, depresión, soledad, tristeza, preocupación/ansiedad por el futuro y dificultad para dormir), "espiritualidad" (3 ítems: esperanza en el más allá, expiación por el pasado y consuelo religioso), "síntomas digestivos" (3 ítems: náuseas, vómitos y pérdida apetito), "familia y economía" (4 ítems: rol como miembro de la familia, preocupación por la familia, problema de vida familiar y carga económica), "actividad diaria" (3 ítems: dependencia en la cama, actividad limitada de la vida diaria y debilidad /cansancio), "atención social" (3 ítems: satisfacción con el cuidado físico, ambiente confortable y satisfacción con el personal médico), "valor por la vida" (4 ítems: actitud positiva, vida significativa, satisfacción con la propia vida y preparación para la muerte), "autoestima" (3 ítems: toma de decisiones independiente, satisfacción con el apoyo familiar y terminar con la vida), "síntomas respiratorios" (2 ítems: tos y disnea), "estreñimiento" e "imagen corporal" (2 ítems: insatisfacción con cambio de apariencia e hinchazón).

Tanto la consistencia interna como la fiabilidad fueron clasificadas como "inadecuadas", resultando un alfa de Cronbach entre 0.41 y 0.93 y coeficientes de correlación test-retest de 0.29-0.83. Finalmente, para la prueba de hipótesis la calificación fue "muy buena", y HQLS se correlacionó significativamente con el dominio "funcionamiento físico" de EORTC QLQ-C30 y con el dominio "síntomas físicos" del MQLS.

3.1.15 Perform Questionnaire (PQ)

Algunas de las propiedades psicométricas de PQ se evaluaron en un único estudio (76). La calidad metodológica para la consistencia interna fue calificada como "muy buena", obteniendo un alfa de Cronbach de 0.94. La fiabilidad test-retest fue catalogada como "inadecuada", con un ICC de 0.83. Para la prueba de

hipótesis la calificación resultó "muy buena", la comparación de las puntuaciones de PQ con las otras medidas de salud (Functional Assessment of Cancer Therapy-Fatigue (FACT-F), Nottingham Health Profile (NHP) y Karnofsky index) reveló que las correlaciones más fuertes se produjeron entre los puntajes de Perform y FACT-F (total = 0.80; $p < 0.05$; dimensiones = 0.68-0.75; $p < 0.05$).

3.1.16 Symptom Assessment Scale (SAS)

SAS se evaluó en un solo estudio (77). La consistencia interna de las propiedades psicométricas fue evaluada y clasificada según su calidad metodológica como "inadecuada", con valores para el alfa de Cronbach entre 0.64 y 0.92. La fiabilidad test-retest se clasificó como "dudosa", con un ICC entre 0.84 y 0.92. Finalmente, la prueba de hipótesis se catalogó como "muy buena", y para esta se encontraron diferencias significativas entre el diagnóstico de cáncer y las puntuaciones de síntomas intestinales ($F = 2.334$, $p < 0.05$), las puntuaciones de síntomas respiratorios ($F = 7.540$, $p < 0.001$) y el dolor ($F = 4.326$, $p < 0.001$).

3.1.17 Three-Levels-of-Needs Questionnaire (3LNQ)

3LNQ estuvo presente en un único artículo (78). La propiedad psicométrica de validez de contenido fue evaluada y clasificada como "inadecuada". Para la prueba de hipótesis la calificación otorgada fue "muy buena", registrándose valores de kappa entre 0.05 y 0.83.

3.1.18 Four-Item Fatigue Scale (FIFS)

FIFS se evaluó únicamente para un estudio (50) y solo se midieron 2 propiedades psicométricas: prueba de hipótesis y capacidad de respuesta. La prueba de hipótesis se calificó como "adecuada", FIFS y BFI se correlacionaron con ciertos ítems de ICD-10, y la correlación de Spearman entre FIFS y BFI fue de 0.69 el día 1 y de 0.73 una semana después ($p < 0.05$).

3.1.19 European Organization for Research and Treatment of Cancer Quality of Life 12 items Questionnaire (EORTC QLQ-FA12)

Para EORTC QLQ-FA12 algunas de las propiedades psicométricas se evaluaron solamente en un estudio (79). La propiedad psicométrica de validez de contenido fue evaluada y calificada como "inadecuada". La validez estructural se catalogó como "adecuada", y se realizó un CFA en el que los 3 factores resultantes fueron "físico", "emocional" y "cognitivo". La puntuación de consistencia interna fue "inadecuada", con valores de alfa de Cronbach > 0.70 . Finalmente, la calificación para fiabilidad test-retest fue "dudosa" y obtuvo valores de ICC entre 0.905 y 0.921.

4. DISCUSIÓN

Se trata de la primera revisión sistemática en la que se identifican los instrumentos para medir FRC en CP y en la que se evalúan las propiedades psicométricas y la calidad metodológica de los instrumentos basándonos en las normas COSMIN (17,22). Se identificaron 19 instrumentos diseñados específicamente para evaluar FRC en CP a través de la inclusión de 57 estudios.

Diferentes factores como el ámbito y la población donde se utilizará el instrumento, sus dimensiones, el número de ítems y la evidencia mostrada en la evaluación de cada propiedad psicométrica, deben ser tenidos en cuenta por clínicos e investigadores para decidir qué instrumento es el más adecuado para medir este constructo. Además, se debe considerar el idioma y la cultura de la versión original del instrumento a emplear.

No hubo ningún instrumento para medir la FRC en CP para el cual se evaluaron todas las propiedades psicométricas, siendo error de medición la única propiedad que no estuvo presente en ninguno de los 19 instrumentos incluidos para dicha revisión. Las propiedades evaluadas para la mayoría de los estudios fueron consistencia interna, validez transcultural, prueba de hipótesis y sensibilidad al cambio.

Según la metodología COSMIN, la validez de contenido, la validez estructural y la consistencia interna son las propiedades psicométricas más importantes (17,21).

La calidad metodológica de los estudios para cada una de las propiedades psicométricas evaluadas fue muy variada, con excepción de la validez de contenido, en la que la calificación obtenida fue "inadecuada" para los 10 instrumentos en los que estuvo presente, debido principalmente a que el método usado no era apropiado para pacientes o profesionales a quienes no se les preguntó sobre la relevancia, exhaustividad y/o comprensibilidad de las instrucciones de la medida de resultado informada por el paciente (PROM), los ítems, las opciones de respuesta o el período de recuperación.

La validez estructural fue calificada como "inadecuada" o "dudosa" en 3 de los 10 instrumentos: EORTC QLQ-C30 (27,28), CFS (47) y BFI (48,49). En consecuencia, en la mayoría de los casos, el tamaño de la muestra incluida en el análisis no fue el adecuado o, en menor medida, la elección del modelo no se ajustó bien a la pregunta de investigación y/o hubo otros fallos importantes en el diseño o métodos estadísticos del estudio. Por otra parte, la consistencia interna fue catalogada como "inadecuada" en 7 de los instrumentos en los que se probó, debido principalmente a que en los puntajes basados en la teoría de respuesta al ítem (IRT) no se calculó el error estándar de theta (SE (θ)) o el coeficiente de confiabilidad del valor del rasgo latente estimado.

Para la validez de contenido, la validez estructural y la consistencia interna; la calidad metodológica, la calidad de la evidencia y la calidad de los resultados obtuvieron valoraciones muy variadas. Adicionalmente, para la validez de contenido la calidad de la evidencia fue "baja" o "muy baja" en los 10 instrumentos en los que estuvo presente, de lo que se deduce que la estimación de la propiedad de medida es limitada y, por tanto, la propiedad de medida real puede ser o es sustancialmente diferente de la estimación de la propiedad de medición. En cuanto a la calidad de los resultados, la calificación fue de "+ -" para 8 de los 10 instrumentos, indicativo de que cada uno de los ítems incluidos son inconsistentes o indeterminados con respecto al constructo que se pretendía medir.

La validez estructural estuvo presente en 10 de los 19 instrumentos y la calidad de la evidencia fue "alta o moderada" para 8 de estos instrumentos (EORTC QLQ-C30 (27,28), ESAS (33,44,45), PQLI (52), PNPC (55), EORTC QLQ-C15-PAL (62-64), FACIT-PAL (73,74), HQLS (80) y EORTC QLQ-FA12 (79)), indicativo de la certeza o confianza moderada en que la propiedad de medición

verdadera se encuentra cerca de la estimación de la propiedad de medición, pero al mismo tiempo también existe una posibilidad de que sea sustancialmente diferente. La calidad de los resultados también fue "?" para 8 de los 10 instrumentos, principalmente por la falta de información sobre el ajuste del modelo en los estudios donde estuvo presente.

En varias de estas herramientas se realizó un EFA, resultando una solución factorial con un rango mínimo de entre un factor para BFI (48,49) y un rango máximo de 11 factores para HQLS (80). Para PNPC (55), EORTC QLQ-C15-PAL (62,63), FACIT-PAL (73) y EORTC QLQ-FA12 (79) se ejecutó un CFA y el rango para la solución factorial osciló entre 2 factores para el instrumento PNPC (55), y 5 factores para EORTC QLQ-C15-PAL (63) y FACIT-PAL (73). También se llevó a cabo un análisis Rash (modelo Rash de crédito parcial) para el instrumento ESAS (45).

Para concluir, la consistencia interna obtuvo una gran variedad de posibles resultados para los 16 instrumentos en los que se calculó. La mayoría de estos instrumentos fueron calificados como "altos" (EORTC QLQ-C30 (25-27,29), CFS (47), BFI (48,49), CAMPAS-R (51), PQLI (52), MFI-20 (70) y PQ (76)) y para el resto fue desigual, obteniéndose una calificación "moderada" o "baja". Para la calidad de los resultados, también existieron diferencias entre las herramientas evaluadas, resultando "+" (alfa de Cronbach ≥ 0.70) y "-" (alfa de Cronbach < 0.70). Es importante destacar que EORTC QLQ-C30 se calificó como "+" en todos los estudios, mientras que EORTC QLQ-C15-PAL se calificó como "-", a pesar de ser la misma escala pero con varios ítems diferentes. Una puntuación inferior a alfa de Cronbach > 0.70 podría influir en la decisión de la herramienta a utilizar para evaluar la FRC en CP, ya que la consistencia interna es una de las propiedades psicométricas más significativas, por lo que aquellas herramientas que puntuaron como "-" deberían ser evaluadas por sus otras características para ser utilizadas en dicha población.

El instrumento ESAS fue el cuestionario incluido en un mayor número de estudios (15 estudios en total) y el segundo, junto con PNPC, en el que más características psicométricas se evaluaron (7 en total), por detrás de EORTC QLQ-C15-PAL en el que fueron evaluadas 8 en total, y este además fue el segundo con mayor presencia en cuanto a número de estudios (14 en total). Para estos 3 instrumentos (ESAS, PNPC y EORTC QLQ-C15-PAL) la calidad metodológica de

las 3 propiedades más importantes (validez de contenido, validez estructural y consistencia interna) fue idéntica ("inadecuada", "adecuada" e "inadecuada", respectivamente). En cuanto a la calidad de los resultados, la mejor puntuación fue para PNPC (53-55), donde la validez de contenido fue "+ -", la validez estructural "+" y la consistencia interna "-". Finalmente, para la calidad de la evidencia, al igual que para la calidad metodológica, la mejor puntuación se distribuyó entre ESAS y PNPC, resultando calificaciones de "baja", "moderada" y "moderada" para las tres propiedades psicométricas mencionadas anteriormente, respectivamente.

HQLI (24), CFS (47), CAMPAS-R (51), PQLI (52), MFI-20 (70), HQLS (80), PQ (76), SAS (77), 3LNQ (78), FIFS (50) y EORTC QLQ-FA12 (79) solo se evaluaron en un artículo y, por tanto, se necesitan estudios adicionales para evaluar las propiedades psicométricas de estos instrumentos. Para PQLI (52), HQLS (80) y EORTC QLQ-FA12 (79), cabe señalar que aunque se incluyeron en un único estudio, se evaluaron la validez de contenido, la validez estructural y la consistencia interna para estos 3 instrumentos.

Es importante destacar la congruencia encontrada en la herramienta EORTC QLQ en relación con estudios previos. Esta herramienta ha sido recomendada en revisiones sistemáticas anteriores para evaluar la fatiga en personas con cáncer general (81,82), al igual que en el presente estudio. Por el contrario, la versión de las escalas es diferente, ya que en cáncer paliativo la herramienta más recomendada es la versión de 15 ítems (EORTC QLQ-C15-PAL), mientras que en el cáncer en general los autores recomiendan la versión de 30 ítems (EORTC QLQ-C30).

Si compramos los resultados de la presente revisión sistemática con investigaciones previas, BFI también se incluyó como una herramienta válida para medir la fatiga en el cáncer en general (81-83) y en diferentes poblaciones (84). En contraste, en el presente estudio, BFI no fue una herramienta recomendada para ser utilizada en cáncer paliativo de acuerdo a las propiedades psicométricas. Esto podría explicarse por el hecho de que BFI es una herramienta autoadministrada que puede ser muy difícil de utilizar en pacientes oncológicos paliativos, por lo que su uso en esta población puede verse limitado.

Otras herramientas de valoración de fatiga en cáncer general también han sido incluidas en la presente revisión sistemática, como FACIT-PAL y MFI-20,

aunque estas, finalmente, no han sido recomendadas para ser empleadas en pacientes con cáncer. Por lo que estas herramientas para evaluar fatiga deben ser específicas y válidas para cada población.

Otro aspecto a tener en cuenta, es la edad de los pacientes a evaluar, ya que las herramientas que se pueden emplear varían en función de esta. Los instrumentos de evaluación incluidos en la presente revisión sistemática no aparecen en ningún estudio similar centrado en niños y/o adolescentes (85), por lo que sería conveniente validar herramientas que midan la FRC en CP en niños, dado que los instrumentos incluidos en el presente estudio pueden no ser adecuados para este tipo de población.

Podemos concluir que ESAS, PNPC y EORTC QLQ-C15-PAL fueron prácticamente los cuestionarios con mayor presencia en cuanto a número de estudios, con más propiedades psicométricas evaluadas y de más importancia (validez de contenido, validez estructural y consistencia interna), resultando estos 3 cuestionarios los más ampliamente validados y más recomendados para su uso en la evaluación de FRC en CP.

Para esta evaluación también se debe incluir PQLI (52), ya que a pesar de haber sido utilizado en un solo estudio, presentó las mejores puntuaciones en cuanto a calidad metodológica y calidad de resultados para las propiedades psicométricas más importantes de todos los instrumentos presentes en dicha revisión sistemática (validez de contenido, validez estructural y consistencia interna), y también evaluó 6 de las 9 propiedades de medición (validez de contenido, validez estructural, consistencia interna, validez transcultural, prueba de hipótesis y sensibilidad al cambio), siempre teniendo en cuenta la necesidad de más estudios adicionales donde se evalúen dichas propiedades psicométricas para este instrumento.

En cuanto a HQLS (80) y EORTC QLQ-FA12 (79) quedarían descartados, ya que HQLS (80) aunque también valoró 6 de las 9 propiedades psicométricas, incluidas las 3 más importantes (validez de contenido, validez estructural, consistencia interna, fiabilidad, prueba de hipótesis y sensibilidad al cambio), no obtuvo buenos resultados en cuanto a calidad metodológica, calidad de resultados y calidad de la evidencia, además de estar presente en un solo estudio.

Por su parte, EORTC QLQ-FA12 (79), si bien obtuvo las mejores puntuaciones en cuanto a la calidad de los resultados para las 3 propiedades

psicométricas más significativas, solamente evaluó 4 de las 9 propiedades psicométricas posibles (validez de contenido, validez estructural, consistencia interna y fiabilidad) y también estuvo presente en un solo estudio.

4.1 Fortalezas y limitaciones

Hasta donde sabemos, esta es la revisión sistemática más actualizada de los instrumentos que evalúan FRC en CP. Esta revisión se realizó siguiendo el dictamen de COSMIN, que aseguró que se utilizó un método adecuado, siguiendo las recomendaciones de los expertos. Es posible que no se hayan identificado todas las herramientas para la evaluación de la fatiga en personas con cáncer en CP, ya que únicamente se seleccionaron artículos publicados en inglés y español.

4.2 Implicaciones prácticas

La evaluación de la FRC en CP puede ser útil para facilitar el cribado, monitorizar la evolución y orientar las decisiones de tratamiento. Se dispone de varios cuestionarios validados para evaluar la fatiga en CP, siendo fundamental elegir el instrumento más adecuado para evaluar este constructo. Esta revisión sistemática proporciona una evaluación exhaustiva de la evidencia actual de los instrumentos que evalúan la fatiga en CP, y los resultados obtenidos van a servir a médicos e investigadores de referencia sobre la calidad de dichos instrumentos. Las herramientas más recomendadas según sus propiedades psicométricas son ESAS, PNPC, EORT QLQ-C15-PAL y PQLI para la práctica clínica y la investigación, pero es importante considerar otras características propias de cada herramienta como la población que se evalúa, el idioma, la cantidad de ítems o el tiempo de aplicación de dicha herramienta.

5. CONCLUSIONES

Se han desarrollado diversos instrumentos para la evaluación de la FRC en CP. La evaluación de este constructo es de gran utilidad, tanto en la práctica clínica como en la investigación, requiriendo nuevos conocimientos en esta área.

La evaluación de la fatiga en esta población es crucial, ya que los pacientes con cáncer avanzado experimentan múltiples síntomas, entre ellos fatiga, siendo este uno de los más frecuentes e importantes en el desenlace de la enfermedad, y en muchos casos descuidado por el personal sanitario al valorar su gravedad y poner tratamiento para atenuarlo. ESAS, PNPC, EORTC QLQC15-PAL y PQLI son los instrumentos recomendados para evaluar la FRC en CP, teniendo en cuenta que es necesario un mayor número de estudios con una alta calidad metodológica en los cuales sean analizadas todas las propiedades psicométricas de los instrumentos ya existentes, así como la adaptación de estos instrumentos a otras culturas.

REFERENCIAS BIBLIOGRÁFICAS

1. Sato A, Matsubayashi K, Morishima T, Nakata K, Kawakami K, Miyashiro I. Increasing trends in the prevalence of prior cancer in newly diagnosed lung, stomach, colorectal, breast, cervical, and corpus uterine cancer patients: a population-based study. *BMC cancer*. 2021;21(1):1-14.
2. Bruera E, Yennurajalingam S. Palliative care in advanced cancer patients: how and when? *The oncologist*. 2012;17(2):267.
3. Fulton JJ, LeBlanc TW, Cutson TM, Porter Starr KN, Kamal A, Ramos K, et al. Integrated outpatient palliative care for patients with advanced cancer: a systematic review and meta-analysis. *Palliative medicine*. 2019;33(2):123-34.
4. Rangachari D, Smith TJ. Integrating palliative care in oncology: the oncologist as a primary palliative care provider. *Cancer journal (Sudbury, Mass)*. 2013;19(5):373.
5. Augustussen M, Sjøgren P, Timm H, Hounsgaard L, Pedersen ML. Symptoms and health-related quality of life in patients with advanced cancer - A population-based study in Greenland. *Eur J Oncol Nurs*. junio de 2017;28:92-7.
6. Teunissen SC, Wesker W, Kruitwagen C, de Haes HC, Voest EE, de Graeff A. Symptom prevalence in patients with incurable cancer: a systematic review. *Journal of pain and symptom management*. 2007;34(1):94-104.
7. Curt GA. The impact of fatigue on patients with cancer: overview of FATIGUE 1 and 2. *The Oncologist*. 2000;5:9-12.

8. Vilchynska T, Beard B. Cancer-related fatigue in palliative care: a global perspective. *International journal of palliative nursing*. 2016;22(5):244-52.
9. Bower JE. Cancer-related fatigue—mechanisms, risk factors, and treatments. *Nature reviews Clinical oncology*. 2014;11(10):597.
10. Kirshbaum M. Cancer-related fatigue: a review of nursing interventions. *British Journal of Community Nursing*. 2010;15(5):214-9.
11. Vilchynska T, Beard B. Cancer-related fatigue in palliative care: a global perspective. *Int J Palliat Nurs*. 2 de mayo de 2016;22(5):244-52.
12. Curt GA, Breitbart W, Cella D, Groopman JE, Horning SJ, Itri LM, et al. Impact of cancer-related fatigue on the lives of patients: new findings from the Fatigue Coalition. *The oncologist*. 2000;5(5):353-60.
13. Minton O, Stone P. A systematic review of the scales used for the measurement of cancer-related fatigue (CRF). *Annals of oncology*. 2009;20(1):17-25.
14. Al Maqbali M, Hughes C, Gracey J, Rankin J, Dunwoody L, Hacker E. Quality assessment criteria: psychometric properties of measurement tools for cancer related fatigue. *Acta Oncologica*. 2019;58(9):1286-97.
15. Seyidova-Khoshknabi D, Davis MP, Walsh D. A systematic review of cancer-related fatigue measurement questionnaires. *American Journal of Hospice and Palliative Medicine®*. 2011;28(2):119-29.

16. Prinsen CA, Vohra S, Rose MR, Boers M, Tugwell P, Clarke M, et al. How to select outcome measurement instruments for outcomes included in a “Core Outcome Set”—a practical guideline. *Trials*. 2016;17(1):1-10.
17. Prinsen CA, Mokkink LB, Bouter LM, Alonso J, Patrick DL, De Vet HC, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research*. 2018;27(5):1147-57.
18. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic reviews*. 2015;4(1):1-9.
19. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS peer review of electronic search strategies: 2015 guideline statement. *Journal of clinical epidemiology*. 2016;75:40-6.
20. Terwee CB, Jansma EP, Riphagen II, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Quality of Life Research*. 2009;18(8):1115-23.
21. Mokkink LB, De Vet HC, Prinsen CA, Patrick DL, Alonso J, Bouter LM, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Quality of Life Research*. 2018;27(5):1171-9.
22. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of clinical epidemiology*. 2007;60(1):34-42.

23. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *Bmj*. 2008;336(7650):924-6.
24. McMillan SC, Mahon M. Measuring quality of life in hospice patients using a newly developed Hospice Quality of Life Index. *Quality of Life Research*. 1994;3(6):437-47.
25. Kaasa S, Bjordal K, Aaronson N, Moum T, Wist E, Hagen S, et al. The EORTC core quality of life questionnaire (QLQ-C30): validity and reliability when analysed with patients treated with palliative radiotherapy. *European journal of cancer*. 1995;31(13-14):2260-3.
26. Godoy Fernández MJ, Rojas Tejada AJ, García Puche JL, Cabrera Iboleón J. Fiabilidad y validez de la versión española del EORTC QLQ-C30: medida de la calidad de vida en pacientes oncológicos avanzados. *Revista de Psicología de la Salud*. 1999;11(1):125-39.
27. Mystakidou K, Tsilika E, Parpa E, Kalaidopoulou O, Smyrniotis V, Vlahos L. The EORTC core quality of life questionnaire (QLQ-C30, version 3.0) in terminally ill cancer patients under palliative care: Validity and reliability in a Hellenic sample. *International Journal of Cancer*. 2001;94(1):135-9.
28. Knobel H, Loge JH, Brenne E, Fayers P, Hjermstad MJ, Kaasa S. The validity of EORTC QLQ-C30 fatigue scale in advanced cancer patients and cancer survivors. *Palliative medicine*. 2003;17(8):664-72.

29. Nicklasson M, Bergman B. Validity, reliability and clinical relevance of EORTC QLQ-C30 and LC13 in patients with chest malignancies in a palliative setting. *Quality of Life Research*. 2007;16(6):1019-28.
30. Sterkenburg CA, King B, Woodward CA. A reliability and validity study of the McMaster Quality of Life Scale (MQLS) for a palliative population. *Journal of Palliative Care*. 1996;12(1):18-25.
31. Kim KU, Yoon S-J, Lee J, Ahn H-S, Park H-J, Lee S-I, et al. Validation of the Korean version of the McMaster Quality of Life Scale in terminal cancer patients. *Journal of palliative care*. 2006;22(1):40-5.
32. Philip J, Smith WB, Craft P, Lickiss N. Concurrent validity of the modified Edmonton symptom assessment system with the Rotterdam symptom checklist and the brief pain inventory. *Supportive Care in Cancer*. 1998;6(6):539-41.
33. Carvajal A, Centeno C, Watson R, Bruera E. A comprehensive study of psychometric properties of the Edmonton Symptom Assessment System (ESAS) in Spanish advanced cancer patients. *European Journal of Cancer*. 2011;47(12):1863-72.
34. Chinda M, Jaturapatporn D, Kirshen AJ, Udomsubpayakul U. Reliability and validity of a Thai version of the Edmonton Symptom Assessment Scale (ESAS-Thai). *Journal of pain and symptom management*. 2011;42(6):954-60.
35. Claessens P, Menten J, Schotsmans P, Broeckaert B. Development and validation of a modified version of the Edmonton Symptom Assessment

- Scale in a Flemish palliative care population. *American Journal of Hospice and Palliative Medicine*®. 2011;28(7):475-82.
36. Carvajal A, Hribernik N, Duarte E, Sanz-Rubiales A, Centeno C. The Spanish version of the Edmonton Symptom Assessment System-revised (ESAS-r): first psychometric analysis involving patients with advanced cancer. *Journal of pain and symptom management*. 2013;45(1):129-36.
 37. Kwon JH, Nam S-H, Koh S, Hong YS, Lee KH, Shin S-W, et al. Validation of the Edmonton Symptom Assessment System in Korean patients with cancer. *Journal of pain and symptom management*. 2013;46(6):947-56.
 38. Hannon B, Dyck M, Pope A, Swami N, Banerjee S, Mak E, et al. Modified Edmonton Symptom Assessment System including constipation and sleep: validation in outpatients with cancer. *Journal of pain and symptom management*. 2015;49(5):945-52.
 39. Paiva CE, Manfredini LL, Paiva BSR, Hui D, Bruera E. The Brazilian version of the Edmonton Symptom Assessment System (ESAS) is a feasible, valid and reliable instrument for the measurement of symptoms in advanced cancer patients. *PLoS One*. 2015;10(7):e0132073.
 40. Yokomichi N, Morita T, Nitto A, Takahashi N, Miyamoto S, Nishie H, et al. Validation of the Japanese Version of the Edmonton Symptom Assessment System–Revised. *Journal of pain and symptom management*. 2015;50(5):718-23.

41. Gretarsdottir H, Fridriksdottir N, Gunnarsdottir S. Psychometric properties of the Icelandic version of the revised Edmonton Symptom Assessment Scale. *Journal of pain and symptom management*. 2016;51(1):133-7.
42. Nekolaichuk C, Huot A, Gratton V, Bush SH, Tarumi Y, Watanabe SM. Development of a French Version of the Edmonton Symptom Assessment System-Revised: A Pilot Study of Palliative Care Patients' Perspectives. *Journal of Palliative Medicine*. 2017;20(9):966-76.
43. Pautex S, Vayne-Bossert P, Bernard M, Beauverd M, Cantin B, Mazzocato C, et al. Validation of the French version of the Edmonton symptom assessment system. *Journal of pain and symptom management*. 2017;54(5):721-6.
44. Mori M, Morita T, Yokomichi N, Nitto A, Takahashi N, Miyamoto S, et al. Validation of the Edmonton Symptom Assessment System: Ascites Modification. *Journal of pain and symptom management*. 2018;55(6):1557-63.
45. Sprague E, Siegert RJ, Medvedev O, Roberts MH. Rasch analysis of the Edmonton symptom assessment system. *Journal of pain and symptom management*. 2018;55(5):1356-63.
46. Tricou C, Ruer M, Nekolaichuk C, Sigurdardottir K, Perceau-Chambard E, Bruera E, et al. Transcultural Validation of the French Version of the Modified Edmonton Symptom Assessment Scale: The ESAS12-F. *Journal of palliative medicine*. 2019;22(12):1553-60.
47. Okuyama T, Akechi T, Kugaya A, Okamura H, Shima Y, Maruguchi M, et al. Development and validation of the cancer fatigue scale: a brief, three-

- dimensional, self-rating scale for assessment of fatigue in cancer patients. *Journal of pain and symptom management*. 2000;19(1):5-14.
48. Okuyama T, Wang XS, Akechi T, Mendoza TR, Hosaka T, Cleeland CS, et al. Validation study of the Japanese version of the brief fatigue inventory. *Journal of pain and symptom management*. 2003;25(2):106-17.
49. Mystakidou K, Tsilika E, Parpa E, Mendoza TR, Pisteovou-Gombaki K, Vlahos L, et al. Psychometric properties of the brief fatigue inventory in Greek patients with advanced cancer. *Journal of pain and symptom management*. 2008;36(4):367-73.
50. Davis MP, Khoshknabi D, Walsh D, Lagman R, Karafa MT, Aktas A, et al. Four-item fatigue screen: replacing the brief fatigue index. *American Journal of Hospice and Palliative Medicine®*. 2013;30(7):652-6.
51. Ewing G, Todd C, Rogers M, Barclay S, McCabe J, Martin A. Validation of a symptom measure suitable for use among palliative care patients in the community: CAMPAS-R. *Journal of pain and symptom management*. 2004;27(4):287-99.
52. Mystakidou K, Tsilika E, Kouloulis V, Parpa E, Katsouda E, Kouvaris J, et al. The " palliative care quality of life instrument (PQLI)" in terminal cancer patients. *Health and Quality of Life Outcomes*. 2004;2(1):8.
53. Osse BH, Vernooij MJ, Schadé E, Grol RP. Towards a new clinical tool for needs assessment in the palliative care of cancer patients: the PNPC instrument. *Journal of pain and symptom management*. 2004;28(4):329-41.

54. Osse BH, Vernooij-Dassen MJ, Schadé E, Grol RP. A practical instrument to explore patients' needs in palliative care: the Problems and Needs in Palliative Care questionnaire—short version. *Palliative Medicine*. 2007;21(5):391-9.
55. Wang T, Molassiotis A, Chung BPM, Tan J-Y. Psychometric assessment of the Chinese version of the Problems and Needs in Palliative Care questionnaire-short version in advanced cancer patients. *BMC palliative care*. 2019;18(1):1-12.
56. Groenvold M, Petersen MA, Aaronson NK, Arraras JI, Blazeby JM, Bottomley A, et al. The development of the EORTC QLQ-C15-PAL: a shortened questionnaire for cancer patients in palliative care. *European Journal of Cancer*. 2006;42(1):55-64.
57. Suárez-del-Real Y, Allende-Pérez S, Alférez-Mancera A, Rodríguez RB, Jiménez-Toxtle S, Mohar A, et al. Validation of the Mexican-Spanish version of the EORTC QLQ-C15-PAL questionnaire for the evaluation of health-related quality of life in patients on palliative care. *Psycho-Oncology*. 2011;20(8):889-96.
58. Miyazaki K, Suzukamo Y, Shimosuma K, Nakayama T. Verification of the psychometric properties of the Japanese version of the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 15 palliative (EORTCQLQ-C15-PAL). *Quality of Life Research*. 2012;21(2):335-40.

59. Shin DW, Choi JE, Miyashita M, Choi JY, Kang J, Baik YJ, et al. Cross-cultural application of the Korean version of the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire-Core 15-palliative care. *Journal of pain and symptom management*. 2011;41(2):478-84.
60. Leppert W, Majkovicz M. Validation of the Polish version of the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire-Core 15-Palliative Care in patients with advanced cancer. *Palliative medicine*. 2013;27(5):470-7.
61. Arraras JI, de la Vega FA, Asin G, Rico M, Zarandona U, Eito C, et al. The EORTC QLQ-C15-PAL questionnaire: validation study for Spanish bone metastases patients. *Quality of Life Research*. 2014;23(3):849-55.
62. Nunes NAH. The quality of life of Brazilian patients in palliative care: validation of the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 15 PAL (EORTC QLQ-C15-PAL). *Supportive Care in Cancer*. 2014;22(6):1595-600.
63. Alawneh A, Yasin H, Khirfan G, Qayas BA, Ammar K, Rimawi D, et al. Psychometric properties of the Arabic version of EORTC QLQ-C15-PAL among cancer patients in Jordan. *Supportive Care in Cancer*. 2016;24(6):2455-62.
64. Miyashita M, Wada M, Morita T, Ishida M, Onishi H, Sasaki Y, et al. Independent validation of the Japanese version of the EORTC QLQ-C15-PAL for patients with advanced cancer. *Journal of Pain and Symptom Management*. 2015;49(5):953-9.

65. Zhang L, Wang N, Zhang J, Liu J, Luo Z, Sun W, et al. Cross-cultural verification of the EORTC QLQ-C15-PAL questionnaire in mainland China. *Palliative medicine*. 2016;30(4):401-8.
66. Ozcelik H, Guzel Y, Sonmez E, Aksoy F, Uslu R. Reliability and validity of the Turkish version of the EORTC QLQ-C15-PAL for patients with advanced cancer. *Palliative & Supportive Care*. 2016;14(6):628.
67. Raman S, Ding K, Chow E, Meyer RM, Nabid A, Chabot P, et al. A prospective study validating the EORTC QLQ-BM22 bone metastases module in patients with painful bone metastases undergoing palliative radiotherapy. *Radiotherapy and Oncology*. 2016;119(2):208-12.
68. Golčić M, Dobrila-Dintinjana R, Golčić G, Pavlović-Ružić I, Stevanović A, Gović-Golčić L. Quality of life in a hospice: a validation of the Croatian version of the EORTC QLQ-C15-PAL. *American Journal of Hospice and Palliative Medicine®*. 2018;35(8):1085-90.
69. Rojas-Concha L, Hansen MB, Petersen MA, Groenvold M. Content validation of the EORTC QLQ-C15-PAL with advanced cancer patients and health care professionals from palliative care services in Chile. *BMC Palliative Care*. 2020;19(1):1-10.
70. Lundh Hagelin C, Wengström Y, Runesdotter S, Johan Fürst C. The psychometric properties of the Swedish Multidimensional Fatigue Inventory MFI-20 in four different populations. *Acta Oncologica*. 2007;46(1):97-104.
71. Blazeby JM, Fayers P, Conroy T, Sezer O, Ramage J, Rees M. Validation of the European Organization for Research and Treatment of Cancer QLQ-

- LMC21 questionnaire for assessment of patient-reported outcomes during treatment of colorectal liver metastases. *British journal of surgery*. 2009;96(3):291-8.
72. Paradowska D, Bereza K, Sanna B, Kucharska E, Tomaszewska IM, Dudkiewicz Z, et al. A prospective study to validate the Polish language version of the European Organisation for Research and Treatment of Cancer (EORTC) Colorectal Liver Metastases (QLQ-LMC21) module. *European Journal of Oncology Nursing*. 2017;29:148-54.
73. Lyons KD, Bakitas M, Hegel MT, Hanscom B, Hull J, Ahles TA. Reliability and validity of the functional assessment of chronic illness therapy-palliative care (FACIT-pal) scale. *Journal of pain and symptom management*. 2009;37(1):23-32.
74. Siegert R, Selman L, Higginson IJ, Ali Z, Powell RA, Namisango E, et al. A psychometric evaluation of the functional assessment of chronic illness therapy-palliative care (FACIT-Pal) scale with palliative care samples in three African countries. *Journal of pain and symptom management*. 2014;48(5):983-91.
75. Shinall MC, Ely EW, Karlekar M, Robbins SG, Chandrasekhar R, Martin SF. Psychometric properties of the FACIT-pal 14 administered in an outpatient palliative care clinic. *American Journal of Hospice and Palliative Medicine*®. 2018;35(10):1292-4.
76. Baró E, Carulla J, Cassinello J, Colomer R, Mata JG, Gascón P, et al. Psychometric properties of the Perform Questionnaire: a brief scale for

- assessing patient perceptions of fatigue in cancer. *Supportive Care in Cancer*. 2011;19(5):657-66.
77. Aoun SM, Monterosso L, Kristjanson LJ, McConigley R. Measuring symptom distress in palliative care: psychometric properties of the Symptom Assessment Scale (SAS). *Journal of palliative medicine*. 2011;14(3):315-21.
 78. Johnsen AT, Petersen MA, Pedersen L, Groenvold M. Development and initial validation of the Three-Levels-of-Needs Questionnaire for self-assessment of palliative needs in patients with cancer. *Journal of pain and symptom management*. 2011;41(6):1025-39.
 79. Weis J, Tomaszewski KA, Hammerlid E, Ignacio Arraras J, Conroy T, Lanceley A, et al. International psychometric validation of an EORTC quality of life module measuring cancer related fatigue (EORTC QLQ-FA12). *JNCI: Journal of the National Cancer Institute*. 2017;109(5).
 80. Kim SH, Choi YS, Lee J, Oh SC, Yeom CH, Lee MA, et al. Reliability and validity of the Hospice Quality of Life Scale for Korean cancer patients. *Journal of pain and Symptom Management*. 2009;37(2):156-67.
 81. Minton O, Stone P. A systematic review of the scales used for the measurement of cancer-related fatigue (CRF). *Ann Oncol*. enero de 2009;20(1):17-25.
 82. Seyidova-Khoshknabi D, Davis MP, Walsh D. Review article: a systematic review of cancer-related fatigue measurement questionnaires. *Am J Hosp Palliat Care*. marzo de 2011;28(2):119-29.

83. Amarsheda S, Bhise AR. Systematic review of cancer-related fatigue instruments in breast cancer patients. *Palliat Support Care*. 5 de mayo de 2021;1-7.
84. Whitehead L. The measurement of fatigue in chronic illness: a systematic review of unidimensional and multidimensional fatigue measures. *J Pain Symptom Manage*. enero de 2009;37(1):107-28.
85. Tomlinson D, Hinds PS, Ethier M-C, Ness KK, Zupanec S, Sung L. Psychometric properties of instruments used to measure fatigue in children and adolescents with cancer: a systematic review. *J Pain Symptom Manage*. enero de 2013;45(1):83-91.

Artículo 4

VALIDATION OF THE ADOLESCENT/ADULT SENSORY PROFILE (AASP) FOR SPAIN

ARTÍCULO ACTUALMENTE EN REVISIÓN EN AUSTRALIAN OCCUPATIONAL THERAPY.

VALIDATION OF THE ADOLESCENT/ADULT SENSORY PROFILE (AASP) FOR SPAIN

RESUMEN

Propósito: Existe la necesidad de validar herramientas de evaluación culturalmente adaptadas que puedan identificar problemas de reactividad sensorial en adolescentes y adultos. Por tanto, el objetivo de este estudio fue validar Adolescent/Adult Sensory Profile (AASP-60 ítems) para población española.

Materiales y métodos: Un total de 806 adultos participaron en este estudio. Se realizó un CFA. Se calcularon los coeficientes α de Cronbach para todos los ítems y para cada uno de los cuadrantes. Se realizó un análisis del ICC para una muestra de 134 participantes con un intervalo de una semana. La validez del criterio convergente se midió evaluando la relación entre los cuadrantes de la AASP-42, AASP-60 y ASQ.

Resultados: El índice de CFA fue satisfactorio para una versión de 42 ítems (AASP-42). α de Cronbach = 0.83 para todos los ítems, el ICC para cada uno de los 4 cuadrantes fue; 0.90, 0.90, 0.90 y 0.83, respectivamente. Se compararon los ítems de la versión original de la AASP-60 con los ítems de la nueva versión adaptada al español de la AASP-42 y los niveles de correlación encontrados fueron altos ($r = 0.88-0.97$; $p < 0.01$).

Conclusiones: La AASP-42 es una herramienta fiable y válida con excelentes/buenas propiedades psicométricas para evaluar el procesamiento sensorial en población adulta española.

Palabras clave: AASP, procesamiento sensorial, validación, población española, propiedades psicométricas.

ABSTRACT

Purpose: There is a need to validate culturally adapted assessment tools that can identify sensory reactivity problems in adolescents and adults. Therefore, the objective of this study was to validate Adolescent/Adult Sensory Profile (AASP-60 items) for the Spanish population.

Materials and methods: A total of 806 adults participated in this study. A confirmatory factor analysis was performed. Cronbach's α coefficients were calculated for all items and for each of the quadrants. An intraclass correlation coefficient analysis was performed for a sample of 134 participants with one week interval. The validity of the convergent criterion was evaluated assessing the relationship between the quadrants of the AASP-42, AASP-60 and ASQ.

Results: Index for confirmatory factor analysis was satisfactory for a version of 42 items (AASP-42). Cronbach's $\alpha = 0.83$ for all items, the ICC for each of the four quadrants was; 0.90, 0.90, 0.90, and 0.83, respectively. The original version of the AASP-60 items was compared with the new version of the Spanish adaptation of the AASP-42 items and the correlation levels found were high ($r = 0.88-0.97$; $p < 0.01$).

Conclusions: The AASP-42 is a reliable and valid tool with excellent/good psychometric properties to assess sensory processing in Spanish adults population.

Keywords: AASP, sensory processing, validation, Spanish population, psychometric properties.

1. INTRODUCCIÓN

El procesamiento sensorial hace referencia al modo por el cual el sistema nervioso es capaz de recibir, modificar, integrar y organizar los estímulos sensoriales incluyendo aquellas respuestas de tipo conductual como réplica a la información sensorial (1). Este ciclo permite a los seres humanos elegir cómo responder ante las demandas ambientales.

Interiormente y a través del entorno externo, la recepción, integración y organización efectivas de la información sensorial favorecen la transformación de esta en señales a las que podemos responder de forma adaptativa (2). Esta información sensorial se emplea para establecer nuestros mapas sensoriales a nivel corporal y ambiental (3), y este proceso es fundamental para un adecuado aprendizaje y funcionamiento diario.

A veces, sin embargo, el procesamiento sensorial puede verse afectado, dando lugar a un conjunto de anomalías en las que la información sensorial no es procesada correctamente, causando dificultades a nivel funcional para el desarrollo de diversas actividades de la vida diaria, dificultades de aprendizaje y problemas de comportamiento (4,5). Además, pueden verse afectadas más de una modalidad sensorial, ya sea a nivel táctil, auditivo, olfativo, gustativo, visual, vestibular o propioceptivo (5-7), lo que dificulta en el individuo el desempeño de las actividades diarias, la interacción social y el juego (8-14), así como la adquisición de habilidades sociales, el desarrollo de la autoestima y la autoconfianza (15), provocando en última instancia limitaciones en la calidad de vida y el bienestar de la persona (16,17).

El cuestionario Adolescent/Adult Sensory Profile de 60 ítems (AASP-60 ítems) fue seleccionado para este estudio porque se administra, puntúa e interpreta los resultados de una forma rápida y fue diseñado particularmente para medir las habilidades de procesamiento sensorial de adolescentes y adultos que apoyan o interfieren con el desempeño funcional (18,19). Ya fue traducido y validado transculturalmente a la población española (20,21), pero existen algunas propiedades psicométricas de dicho cuestionario que aún no han sido evaluadas.

Ante esta situación, existe la necesidad de validar herramientas de evaluación adaptadas culturalmente que puedan identificar problemas de reactividad sensorial en personas adolescentes y adultas. En la actualidad existe

un déficit en cuanto a la existencia de este tipo de instrumentos de evaluación del proceso sensorial culturalmente sensibles, de ahí la importancia del empleo de perfiles sensoriales para determinar los beneficios multisensoriales en el envejecimiento saludable y anormal, puesto que ofrecen una interesante alternativa para que las tareas de detección multisensorial se utilicen como una herramienta de detección rentable (22). Por lo tanto, el objetivo de este estudio fue validar la AASP- 60 ítems para población española.

2. MÉTODO

2.1 Participantes

Los participantes fueron reclutados durante un período de tiempo de 2 meses a través de una plataforma online con la colaboración de los colegios de enfermería a nivel nacional. En este estudio participaron un total de 806 adultos sin patologías previas. La muestra fue un 90.07% de mujeres y un 9.92% de hombres con una edad media de 35.19 años. Los criterios de inclusión para la captación de los participantes fueron: (a) personas entre 18 y 65 años, (b) sin patologías previas, (c) tener conocimientos informáticos básicos para poder acceder al cuestionario y d) saber leer y escribir.

2.2 Diseño y medidas del estudio

2.2.1 Procedimiento

El cuestionario AASP-60 ítems utilizado para el estudio fue el previamente adaptado culturalmente a la lengua española por Gándara-Gafo et al (20). Se trata de un cuestionario autoadministrado. Este fue distribuido a través de una plataforma online por diferentes redes sociales y a través de compañeros de trabajo.

2.2.2 Ética

El presente estudio obtuvo la aprobación del Comité de Ética en Investigación de la Universidad Católica de Murcia. Se aplicaron los principios éticos de la última Declaración de Helsinki y las Normas de Buenas Prácticas Clínicas.

Según la Ley 41/2002 de 14 de noviembre (vigente ley española), que regula la autonomía, derechos y obligaciones de los pacientes en materia de información e historia clínica, así como el derecho a recibir información y al consentimiento informado.

Todos los datos fueron guardados y tratados con respeto a la privacidad del participante y están sujetos a la Ley de Protección de Datos (Ley Orgánica 15/1999), garantizando la total confidencialidad. La información obtenida se utilizó para los objetivos declarados del estudio.

2.2.3 Variables

El AASP (23) es un cuestionario autoadministrado que consta de 60 ítems que evalúan los niveles de procesamiento sensorial para la vida diaria a través de 6 secciones (gusto / olfato, movimiento, visual, táctil, actividad y auditiva). La totalidad de los ítems se centran en 3 áreas principales. La primera área valora la capacidad de regular las entradas sensoriales que afectan a las actividades de la vida diaria (AVD). La segunda evalúa la capacidad de interpretar distintas entradas sensoriales relacionadas con diferentes sistemas sensoriales. La tercera evalúa la capacidad para procesar los estímulos sensoriales y responder ante ellos (24). Se debe indicar la frecuencia de las respuestas a las distintas experiencias sensoriales para una escala de cinco puntos: 1 (casi nunca), 2 (raras veces), 3 (ocasionalmente), 4 (con frecuencia), y 5 (casi siempre). El cuestionario tarda en completarse de 10 a 15 minutos.

Está basado en el modelo de procesamiento sensorial de Dunn (1997) (10), los principios interactivos de los umbrales neurológicos (altos o bajos) y las respuestas de tipo conductual (activas o pasivas). Los umbrales neurológicos hacen referencia al conjunto de estímulos imprescindibles para que un sistema neuronal sea capaz de responder a la información sensorial, mientras que las

respuestas conductuales hacen referencia al modo por el cual una persona reacciona a sus umbrales sensoriales. La AASP divide los resultados en 4 probables cuadrantes en relación a los perfiles de comportamiento (registro bajo Q1, búsqueda sensorial Q2, sensibilidad sensorial Q3 y evitación sensorial Q4).

El Adult Sensory Questionnaire (ASQ) elaborado por Kinnealey y Oliver en 2002 (25), fue diseñado como una herramienta de detección o tamizaje para el diagnóstico de hipersensibilidad o Defensividad Sensorial (DS) en conjunción con una entrevista más a fondo (26), como el Adult Sensory Interview (ADULT-SI). Es un cuestionario autoadministrado que consta de 26 ítems que se deben contestar con verdadero-falso. Una respuesta verdadera se puntúa con 1, una respuesta falsa con 0. Una puntuación igual o mayor a 10 es indicativo de presencia de DS definida, una puntuación de 6-10 indica DS moderada. Se considera una herramienta adecuada para detectar DS en adultos, en conjunción con la entrevista, ya que presenta buenas propiedades psicométricas (27), además de ser sencillo y corto (su tiempo de administración no supera los 10 minutos).

2.3 Análisis estadístico

Se llevó a cabo un análisis descriptivo de las variables demográficas del AASP y del ASQ. Se resumieron las variables medidas y se determinó el nivel del AASP en los participantes. La prueba de Kaiser-Meyer-Olkin (KMO) evaluó la adecuación del muestreo, mientras que la prueba de esfericidad de Bartlett se empleó para confirmar la redundancia entre las variables; $KMO > 0.5$ se consideró aceptable, junto con valores significativos para la prueba de esfericidad de Bartlett (medida de IBM Kaiser-Meyer Olkin para matriz de correlación de identidad; Estados Unidos, 2011) (28). Se estimó el análisis factorial exploratorio (EFA) con extracción de máxima verosimilitud y rotación varimax para la estructura interna del nuevo cuestionario. Se realizó un análisis factorial confirmatorio (CFA) para determinar la validez estructural del mismo. Para ello, se calcularon las cargas factoriales de las variables y se empleó la estimación de máxima verosimilitud para considerar el ajuste del modelo. Los índices de ajuste del modelo incluyeron chi cuadrado (χ^2), el error cuadrático medio de aproximación (RMSEA) y el índice de ajuste comparativo (CFI). Para RMSEA, los valores ≤ 0.08 reflejaron un ajuste cercano y razonable, mientras que los valores < 0.05 indicaron un ajuste

excelente (29). El CFI varió a lo largo de un continuo de 0 a 1 en el que los valores ≥ 0.80 se consideraron un ajuste satisfactorio y ≥ 0.95 reflejaron un ajuste excelente (30). Solo aquellos ítems cuyas cargas factoriales en el EFA eran > 0.30 , un conocido punto de corte de aceptabilidad (31), se conservaron para análisis posteriores. Se eliminaron los elementos con cargas cruzadas significativas.

Para determinar la consistencia interna de la escala se calcularon los coeficientes α de Cronbach para la totalidad de los ítems y para cada una de los cuadrantes. Se llevó a cabo un análisis del coeficiente de correlación intraclase (ICC) para una muestra de 134 participantes con un intervalo de una semana entre ambas pruebas, para cuantificar la fiabilidad de las mediciones asociadas a las variables cuantitativas continuas. Además, se evaluó el error estándar de medición (SEM) de cada una de las variables incluidas en los ítems de AASP-42 ítems. La validez de criterio convergente fue calculada realizando la relación entre el cuestionario AASP-42 ítems con el cuestionario AASP-60 ítems y el ASQ. El coeficiente de correlación de Pearson fue usado para valorar la validez convergente entre dichos cuestionarios.

Todos los análisis estadísticos se ejecutaron con el paquete estadístico de ciencias sociales para Windows versión 21.0 y SPSS AMOS.

3. RESULTADOS

3.1 Estadística descriptiva

El número total de participantes que completaron el cuestionario fue 806; 9.92% hombres (80) y 90.07% mujeres (726), con una edad media de 35.19 años. Su altura fue de 164.67 cm y su peso de 68.99 kg. La media para cada uno de los cuadrantes del AASP-60 ítems (versión original) resultó: cuadrante 1 "registro bajo" 29.9392, cuadrante 2 "búsqueda sensorial" 48.0575, cuadrante 3 "sensibilidad sensorial" 38.7685 y cuadrante 4 "evitación sensorial" 38.8095, mientras que la media para los cuadrantes del AASP-42 ítems (versión española) fue: cuadrante 1 "registro bajo" 24.6266, cuadrante 2 "búsqueda sensorial" 35.3114, cuadrante 3 "sensibilidad sensorial" 27.5720 y cuadrante 4 "evitación sensorial" 23.9888. La media para el ASQ, cuestionario utilizado en el cálculo de

la validez convergente, fue 10.4467. La media y la DE para cada uno de estos ítems se muestran en la tabla 16.

Tabla 16. Características de los participantes.

	Mean	SD
Edad	35.19	17.762
Altura	164.6725	12.91480
Peso	68.9907	15.67369
Cuadr1 (AASP-60 ítems)	29.9392	7.32030
Cuadr2 (AASP-60 ítems)	48.0575	8.15727
Cuadr3 (AASP-60 ítems)	38.7685	9.50592
Cuadr4 (AASP-60 ítems)	38.8095	9.33143
ASQ	10.4467	5.35128
Cuadr1 (AASP-42 ítems)	24.6266	6.36787
Cuadr2 (AASP-42 ítems)	35.3114	6.50416
Cuadr3 (AASP-42 ítems)	27.5720	7.20725
Cuadr4 (AASP-42 ítems)	23.9888	5.87371

Abreviaturas: (AASP) Adolescent/Adult Sensory Profile; (ASQ) Adult Sensory Questionnaire; (SD) standard deviation.

3.2 Análisis factorial exploratorio (EFA)

La matriz de correlación evaluada preliminarmente mediante la prueba de esfericidad de Bartlett rechazó la hipótesis nula de una matriz identidad (grados de libertad, df 1770; $p < 0.001$), con una medida de adecuación muestral $KMO = 0.86$, indicativo de que la muestra era adecuada (32). Se utilizaron un valor propio > 1 , un valor $> 10\%$ de la varianza y una prueba de criba como criterios para la extracción de factores (28,33). En base a estas condiciones se extrajeron 6 factores, cumpliendo dos de los tres criterios: auto-valor > 1 y curva del scree test (figura 18). Para conocer la carga de los factores se utilizó el método de rotación varimax porque garantiza que no excluimos la posibilidad de expresar un determinado elemento en un factor o en más de un factor.

La carga de los factores en la estructura funcional rotada, con un punto de corte > 0.3 identificó los siguientes ítems en cada factor (tabla 17): Factor 1 (6 ítems): Aud 4, Aud 5, Aud 7, Aud 8, Aud 11 y Vis 10; Factor 2 (8 ítems): Aud 2, Aud 3, Aud 10, Act 2, Act 5, Act 9, Mov 4 y Touch 11; Factor 3 (12 ítems): Mov 2, Mov 6, Vis 1, Vis 3, Aud 1, Aud 9, Taste 2, Taste 4, Taste 8, Act 3, Act 3 y Act 8; Factor 4 (5 ítems): Touch 8, Touch 9, Touch 12, Act 7 y Vis 9; Factor 5 (8 ítems): Mov 1, Mov 3, Mov 7, Mov 8, Vis 4, Vis 5, Vis 6 y Touch 6; Factor 6 (3 ítems): Touch 10, Touch 13 y Vis 7.

Tabla 17. Matriz de carga de factores después de la rotación varimax.

	Factor					
	1	2	3	4	5	6
Aud 5	0.72					
Aud 8	0.67					
Aud 7	0.66					
Aud 11	0.65					
Vis 10	0.65					
Aud 4	0.61					
Aud 10		0.66				
Aud 3		0.54				
Act 5		0.52				
Mov 4		0.49				
Act 2		0.46				
Aud 2		0.45				
Act 9		0.43				
Touch 11		0.41				
Mov 2			0.67			
Mov 6			0.56			
Vis 1			0.49			
Vis 3			0.48			
Aud 9			0.48			
Aud 1			0.47			
Taste 8			0.46			

	Factor					
	1	2	3	4	5	6
Act 8			0.46			
Act 3			0.43			
Act 4			0.40			
Taste 4			0.39			
Taste 2			0.30			
Touch 9				0.71		
Touch 12				0.70		
Act 7				0.58		
Vis 9				0.44		
Touch 8				0.43		
Mov 1					0.55	
Mov 7					0.51	
Vis 6					0.49	
Mov 8					0.46	
Taste 6					0.44	
Vis 5					0.44	
Mov 3					0.40	
Vis 4					0.39	
Touch 10						0.62
Touch 13						0.58
Vis 7						0.43

Aud: auditory processing; **Vis:** visual processing; **Act:** Activity processing; **Mov:** Movement processing; **Taste:** Taste processing; **Touch:** Touch processing.

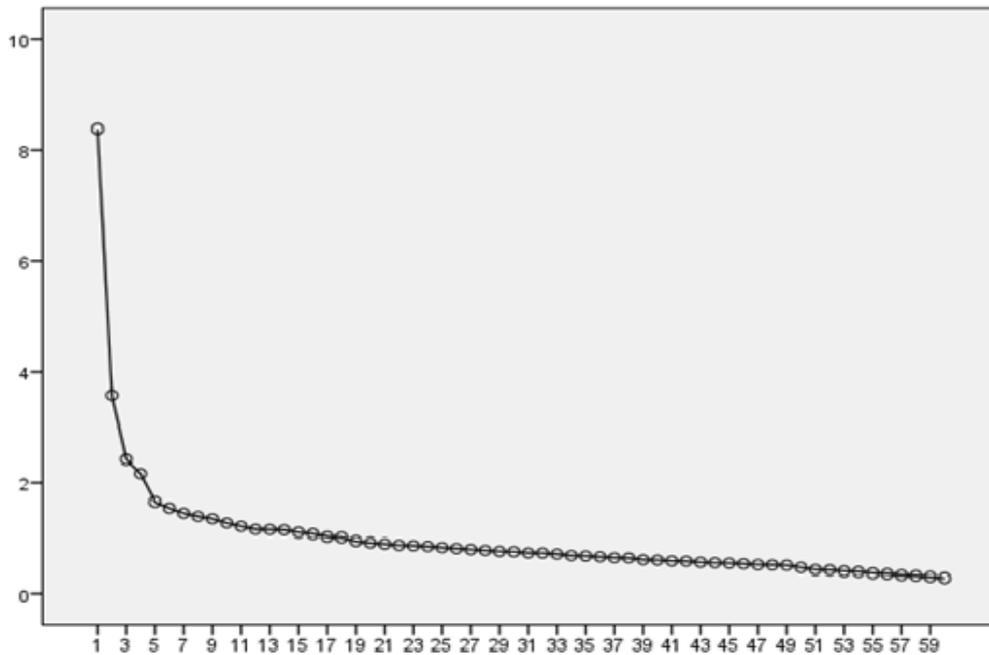


Figura 18. Scree plot (gráfico de pantalla de factores rotados).

3.3 Análisis factorial confirmatorio (CFA)

La mayoría de los índices de ajuste fueron satisfactorios, con CFI = 0.99 y RMSEA = 0.048 para todo el cuestionario AASP-42 ítems. La prueba de chi-cuadrado (χ^2) para los 6 factores fue significativa ($\chi^2 = 2563.1$, $df = 804$, $p > 0.05$).

3.4 Fiabilidad

La consistencia interna, α de Cronbach, para cada uno de los 4 cuadrantes fue; 0.70, 0.69, 0.73 y 0.73, respectivamente. Los valores para el cuestionario completo oscilaron entre $r = 0.816 - 0.831$, resultando para el total de los ítems de $\alpha = 0.83$ (tabla 18).

3.5 Fiabilidad y error de medición

El ICC con un intervalo de una semana entre ambas pruebas para cada uno de los 4 cuadrantes fue; 0.90, 0.90, 0.90 y 0.83, respectivamente. Los valores para cada uno de los ítems variaron entre $r= 0.40-0.92$ (tabla 19). Con estos resultados, el cuestionario AASP parece ser el adecuado para un modelo de 6 factores (figura 19). El error estándar para la medición osciló entre 0.026 - 0.051.

Tabla 18. Estadística descriptiva y consistencia interna para los ítems de la AASP.

AASP ítems	Mean \pm SD	Corrected item-total correlation	Cronbach's alpha if item deleted
Vis10	3.07 \pm 1.41	0.38	0.82
Aud4	2.55 \pm 1.33	0.44	0.82
Aud5	2.93 \pm 1.27	0.56	0.82
Aud7	2.43 \pm 1.37	0.46	0.82
Aud8	2.95 \pm 1.25	0.51	0.82
Aud11	2.46 \pm 1.35	0.43	0.82
Mov4	2.35 \pm 1.22	0.41	0.82
Touch11	3.10 \pm 1.42	0.33	0.82
Act2	2.42 \pm 1.34	0.24	0.83
Act5	2.19 \pm 1.10	0.34	0.82
Act9	2.74 \pm 1.24	0.36	0.82
Aud2	2.80 \pm 1.32	0.43	0.82
Aud3	2.33 \pm 1.13	0.47	0.82
Aud10	2.39 \pm 1.16	0.40	0.82
Sabor2	3.00 \pm 1.33	0.19	0.83
Sabor4	3.50 \pm 1.19	0.15	0.83
Sabor8	3.68 \pm 1.34	0.17	0.83
Mov2	4.04 \pm 1.01	0.13	0.83
Mov6	3.46 \pm 1.25	0.08	0.83
Vis1	2.82 \pm 1.13	0.04	0.83
Vis3	3.30 \pm 1.06	0.15	0.83
Act3	2.82 \pm 1.12	0.27	0.83
Act4	2.88 \pm 1.20	0.07	0.83

AASP ítems	Mean \pm SD	Corrected item-total correlation	Cronbach's alpha if item deleted
Act8	2.38 \pm 1.24	0.05	0.83
Aud1	3.29 \pm 1.30	0.14	0.83
Vis9	2.65 \pm 1.37	0.45	0.82
Touch8	2.76 \pm 1.46	0.29	0.82
Touch9	3.05 \pm 1.29	0.42	0.82
Touch12	2.91 \pm 1.32	0.35	0.82
Touch7	2.62 \pm 1.39	0.41	0.82
Sabor6	1.69 \pm 0.93	0.24	0.83
Mov1	2.96 \pm 1.31	0.21	0.83
Mov3	1.72 \pm 1.12	0.22	0.83
Mov7	1.76 \pm 1.10	0.30	0.82
Mov8	2.49 \pm 1.30	0.39	0.82
Vis4	3.02 \pm 1.34	0.38	0.82
Vis5	2.10 \pm 1.11	0.30	0.82
Vis6	2.70 \pm 1.33	0.42	0.82
Vis7	1.40 \pm 0.74	0.22	0.83
Touch10	1.53 \pm 0.97	0.11	0.83
Touch13	1.30 \pm 0.76	0.09	0.83
Total ítems			0.83
Cuadr1			0.70
Cuadr2			0.69
Cuadr3			0.73
Cuadr4			0.73

Aud: auditory processing; Vis: visual processing; Act: Activity processing; Mov: Movement processing;
Taste: Taste processing; Touch: Touch processing.

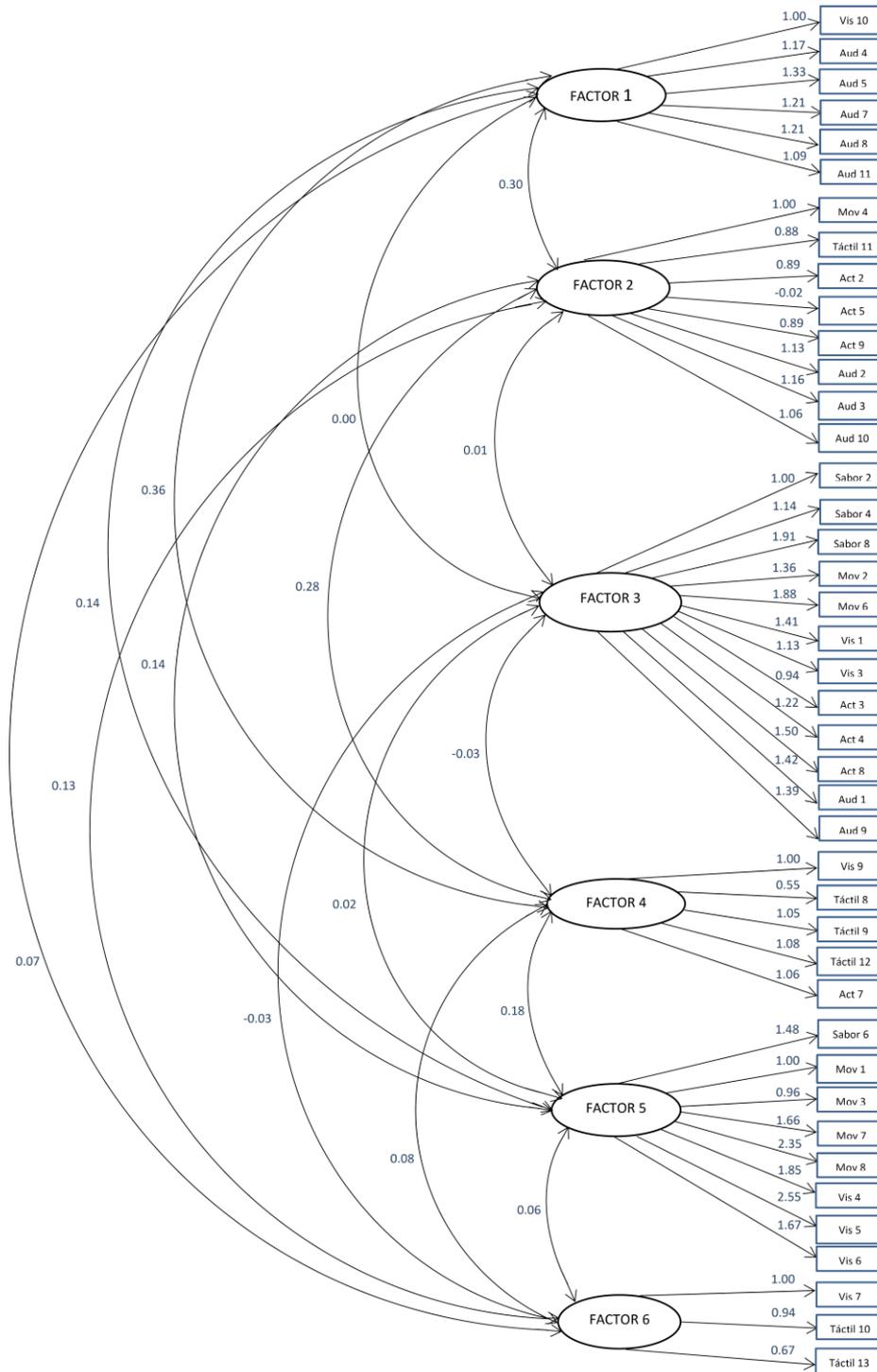


Figura 19. Gráfica análisis factorial confirmatorio (CFA).

Tabla 19. Fiabilidad test-retest y error de medición para el total de la muestra.

AASP ítems	ICC (N=134)	SEM
Vis 10	0.70	0.049
Aud 4	0.77	0.047
Aud 5	0.83	0.045
Aud 7	0.76	0.048
Aud 8	0.73	0.044
Aud 11	0.76	0.048
Mov 4	0.78	0.043
Táctil 11	0.92	0.050
Act 2	0.90	0.047
Act 5	0.81	0.039
Act 9	0.75	0.043
Aud 2	0.82	0.046
Aud 3	0.79	0.039
Aud 10	0.79	0.041
Sabor 2	0.79	0.047
Sabor 4	0.74	0.042
Sabor 8	0.75	0.048
Mov 2	0.88	0.036
Mov 6	0.87	0.044
Vis 1	0.79	0.040
Vis 3	0.88	0.037
Act 3	0.74	0.040
Act 4	0.74	0.042

AASP ítems	ICC (N=134)	SEM
Act 8	0.55	0.044
Aud 1	0.86	0.046
Vis 9	0.82	0.048
Táctil 8	0.65	0.051
Táctil 9	0.74	0.045
Táctil 12	0.74	0.047
Act 7	0.71	0.045
Sabor 6	0.65	0.033
Mov 1	0.89	0.046
Mov 3	0.61	0.039
Mov 7	0.73	0.039
Mov 8	0.78	0.046
Vis 4	0.76	0.047
Vis 5	0.67	0.039
Vis 6	0.76	0.047
Vis 7	0.61	0.026
Táctil 10	0.61	0.034
Táctil 13	0.40	0.027
Cuadr1	0.90	-
Cuadr2	0.90	-
Cuadr3	0.90	-
Cuadr4	0.83	-

Abbreviation: (AASP) Adolescent/Adult Sensory Profile; (ICC) Intraclass Correlation Coefficient; (SEM) standard error of the mean.

3.6 Validez de criterio convergente

En el presente estudio se comparó la versión original del AASP-60 ítems (23) con la nueva versión de la adaptación al español del AASP-42 ítems. En la tabla 20 se muestran las correlaciones de ambos cuestionarios. Los niveles de correlación encontrados fueron altos ($r = 0.88-0.97$; $p < 0.01$), resultando significativos entre ambos cuestionarios para cada uno de sus 4 cuadrantes.

Debido a la falta de un "gold standard" equiparable con el AASP que evalúe los niveles de procesamiento sensorial para la vida diaria, para la validez convergente se comparó el cuestionario AASP-42 ítems con el ASQ (25). Este cuestionario ha sido utilizado en diferentes investigaciones como instrumento de screening para examinar la relación entre DS y diferentes variables en adultos (27,34-36). En la tabla 20 se muestran las correlaciones del cuestionario AASP-42 ítems con el ASQ, resultando niveles de correlación bajos y moderados ($r = -0.28-0.52$; $p < 0.05$), pero significativos entre ambos cuestionarios para cada uno de sus 4 cuadrantes.

Tabla 20. Validez de criterio convergente de AASP- 42 ítems.

	AASP – 42 ítems Cuadr1	AASP – 42 ítems Cuadr2	AASP – 42 ítems Cuadr3	AASP – 42 ítems Cuadr4
AASP – 60 ítems Cuadr1	0.97**			
AASP – 60 ítems Cuadr2		0.94**		
AASP – 60 ítems Cuadr3			0.95**	
AASP – 60 ítems Cuadr4				0.88**
ASQ	0.35*	-0.28*	0.52*	0.41*

Abbreviation: (AASP) Adolescent/Adult Sensory Profile; (ASQ) Adult Sensory Questionnaire.

* $p < 0.05$

** $p < 0.01$

4. DISCUSIÓN

La necesidad de validar herramientas de evaluación adaptadas culturalmente que puedan identificar problemas de reactividad sensorial en personas adolescentes y adultas ha sido estudiada en diferentes poblaciones, debido principalmente a las limitaciones en la calidad de vida y el bienestar que en estas produce con la consiguiente incapacidad para un adecuado desempeño funcional (37-46).

Es de vital importancia utilizar medidas válidas y fiables que se adapten cultural y lingüísticamente a la población objetivo. El análisis de las variables psicométricas de la escala AASP permitió obtener una nueva herramienta para ser usada en la práctica clínica y en investigación para la valoración del procesamiento. Actualmente, los estudios psicométricos con relación a este cuestionario son limitados, y el presente estudio representa un paso crítico hacia la mejora de la medición de las características de procesamiento sensorial en adolescentes y adultos que apoyan o interfieren con el desempeño funcional (18,19), con el fin de planificar una intervención individualizada en función de las características de procesamiento sensorial de cada paciente.

En España, actualmente, las únicas herramientas de evaluación adaptadas culturalmente son el SSP (Short Sensory Profile) (47) y el AASP- 60 ítems (23), en español llamado "Perfil sensorial del adolescente/adulto" (PSAA) (48). Para este cuestionario, AASP- 60 ítems, existen diversas versiones del original adaptadas a diferentes culturas: americana (49), china (50) y árabe (51-54); mientras que la versión corta, SSP, está presente también para distintas poblaciones: americana (55) y polaca (56).

En el presente estudio se valoró la validez estructural del cuestionario AASP llevando a cabo un EFA y posteriormente un CFA. En el EFA, la matriz de correlación rechazó la hipótesis nula de una matriz identidad (grados de libertad, df 1770; $p < 0.001$), con una medida de adecuación muestral $KMO = 0.86$, ofreciendo un resultado notable, lo que implicó que la relación entre las variables fuera alta (32). Se extrajeron 6 factores cumpliendo dos de los tres criterios; autovalor > 1 y curva del scree test (28,33). El CFA realizado posteriormente confirmó esta versión de 6 factores con excelentes ajustes estadísticos. Siguiendo estos

critérios se obtuvieron un total de 6 factores, con un total de 42 ítems (18 ítems fueron eliminados en la matriz de carga).

Este nuevo cuestionario denominado AASP- 42 ítems mantiene los factores del cuestionario original, el cual fue traducido y adaptado culturalmente al español (20). A pesar de que ambos cuestionarios mantienen el mismo número de factores, este nuevo cuestionario ha disminuido a 42 ítems, a diferencia de los 60 ítems del anterior (20). En cambio, esta estructura de 6 factores no es encontrada en todas las versiones de dicho cuestionario, ya que diferentes autores han obtenido versiones con un número diferente de factores: desde un solo factor, el caso de la versión corta adaptada a población polaca de Chojnicka et al (56), a los 4 factores de la versión árabe de Almomani et al (51), o los 9 factores para las versiones americanas de Brown et al (49) y Zacharty et al (55).

Por otro lado, también existió heterogeneidad en el número de ítems total del cuestionario: desde los 125 ítems del SP (Sensory Profile) versión árabe de Kayihan et al. (2015) (52), hasta los 38 ítems del SSP versión americana de Zachary et al. (2018) (55) y los 34 ítems del SSP-2-PL (Polish version of the Short Sensory Profile, 2nd edition) versión polaca de Chojnicka et al. (2019) (56). A pesar de este rango en cuanto al número de ítems, la mayoría de los cuestionarios validados poseen un total de 60 ítems, como el ASP (Adult Sensory Profile) versión americana de Brown et al. (2001) (49), el ASP versión china de Chung et al. (2006) (50), el AASP versión árabe de Almomani et al. (2013) (51), el AASP versión española de Gándara-Gafo et al. (2019) (21) y el AASP versión persa de Zaree (2021) (54). En cambio, el cuestionario validado en el presente estudio para la población española posee un total de 42 ítems, sin que posteriormente encontremos un instrumento similar en cuanto a número de ítems.

En cuanto a la consistencia interna, para nuestro estudio fue calculado α de Cronbach para cada uno de los 4 cuadrantes, resultando 0.70, 0.69, 0.73 y 0.73, respectivamente. Los valores para el cuestionario completo oscilaron entre $r=0.816-0.831$ y para el total de los ítems fue de $r=0.83$, indicativo de la óptima consistencia entre sí de los ítems, equivalente a lo acontecido en otros estudios donde también se utilizó el AASP. Similares resultados con valores α de Cronbach > 0.60 fueron mostrados en diferentes estudios donde fue evaluada la consistencia interna para el cuestionario AASP en su versión original de 60 ítems (21,49,53,54). La consistencia interna de las versiones cortas de dicho cuestionario

también ha sido evaluada, encontrando valores similares a los del presente estudio y a los de las versiones largas (en torno a $\alpha = 0.70$), con valores de 0.70 para el estudio de Zacharty et al. (55) y de $\alpha = 0.72-0.93$ para Chojnicka et al. (56). Es importante destacar que diferentes versiones como la china y la árabe han obtenido valores menores de consistencia interna (50,52).

Los valores de fiabilidad fueron calculados en una muestra de 134 personas con un intervalo de una semana entre ambas pruebas para cada uno de los 4 cuadrantes, resultando 0.90, 0.90, 0.90 y 0.83, respectivamente. Los valores para cada uno de los ítems oscilaron entre $ICC = 0.40 - 0.92$, indicativo de que la concordancia entre dos o más valoraciones cuantitativas (continuas) obtenidas en dos evaluaciones diferentes separadas por un intervalo de tiempo de una semana fueron aceptables, situándose en el caso de los 4 cuadrantes > 0.70 (57-59). Los resultados mostrados en el presente estudio mejoran considerablemente los de la versión española original de 60 ítems, en la cual los valores de $ICC = 0.58-0.95$ para la muestra monolingüe y entre $ICC = 0.46-0.95$ para la muestra bilingüe (21). Dicho cuestionario ha mostrado resultados óptimos en cuanto a ICC en diferentes versiones, obteniendo resultados similares a los del presente estudio, como en el caso de Chung et al. ($ICC = 0.76-0.99$) (50), Almomani et al. ($ICC = 0.80-0.96$) (51), Kayihan et al. en su versión adaptada a población árabe ($ICC > 0.90$) (52), Zaree et al. ($ICC = 0.88-0.94$) (54) y Chojnicka et al. en su versión corta adaptada a población polaca ($ICC = 0.94-0.98$) (56).

El SEM obtuvo valores entre 0.026 - 0.051. Para el resto de versiones de esta escala adaptadas a otras culturas el SEM solo fue calculado en 2 estudios, obteniendo valores mayores que los del presente estudio: el SP versión Turca de Kayihan et al., de 125 ítems, donde el rango obtenido varió entre 1.0 - 2.8 (52) y el AASP versión Persa de Zaree et al., de 60 ítems, donde el rango conseguido osciló entre 1.1 - 2.6 (54).

La validez de criterio convergente fue evaluada comparando la versión original del AASP-60 ítems (23) con la nueva versión de la adaptación al español del AASP-42 ítems, resultando niveles de correlación altos y estadísticamente significativos entre ambos cuestionarios para cada uno de sus 4 cuadrantes ($r = 0.88 - 0.97$; $p < 0.01$), y el AASP-42 ítems con el ASQ reflejando niveles de correlación de bajos a moderados, pero significativos entre ambos cuestionarios para cada uno de sus 4 cuadrantes ($r = -0.28$ a 0.52 ; $p < 0.05$). La validez de criterio

convergente también fue evaluada en el trabajo de Chojnicka et al., donde fue comparada una versión corta del AASP, el SSP-2-PL adaptado a población polaca, con el SCQ (Social Communication Questionnaire), y donde el análisis arrojó correlaciones positivas y estadísticamente significativas con la puntuación total del SCQ ($r = 0.35-0.49$; $p < 0.001$), así como con el SSP-2-PL ($r = 0.47$) (56).

4.1 Fortalezas y limitaciones

A nuestro entender, la fortaleza de este estudio reside en el rigor metodológico de sus análisis psicométricos, además de ser el estudio, de todos los que existen, donde más propiedades psicométricas se han evaluado en relación al AASP. El estudio evaluó una muestra en población española relativamente grande, demostrando que la nueva versión española cuenta con buenos parámetros psicométricos. Por otro lado, el presente estudio tiene varias limitaciones. En primer lugar, existe muy poca evidencia científica en relación a este cuestionario en población adulta, ya que en la mayoría de los artículos donde se ha validado dicho cuestionario y se han evaluado algunas de las propiedades psicométricas solamente se ha tenido en cuenta población menor de 18 años. En segundo lugar, el número desigual de participantes en ambos sexos (hombres versus mujeres) y la inclusión de individuos exclusivamente sanos, lo que puede dar lugar a un sesgo de muestreo.

5. CONCLUSIONES

Los resultados obtenidos en este estudio nos llevan a concluir que la versión española del AASP es adecuada para medir el procesamiento sensorial en esta población, tanto en adolescentes como en adultos y es útil para los terapeutas ocupacionales como herramienta clínica y de investigación. La disponibilidad de un instrumento que mida de una manera fiable los patrones de procesamiento sensorial de los pacientes con disfunción es una ventaja en la práctica clínica, no solo nos va a proporcionar información sistemática sobre los patrones de procesamiento sensorial de los pacientes, sino que también nos va a permitir a los profesionales clínicos formular intervenciones apropiadas para que coincidan con esos mismos patrones. Las excelentes propiedades psicométricas de este

cuestionario mostradas en este estudio en población española indican que podría ser utilizado en futuros estudios de investigación o para futuras intervenciones en personas adolescentes o adultas para medir el procesamiento sensorial, identificar problemas de reactividad sensorial con el fin de planificar una intervención individualizada para cada uno de estos individuos en función de sus características, facilitar información objetiva a las familias sobre el pronóstico o progreso a nivel de procesamiento sensorial de sus hijos/as y para ayudarles a detectar cambios clínicamente significativos.

REFERENCIAS BIBLIOGRÁFICAS

1. Miller Lj, Lane SJ. Toward a consensus in terminology in sensory integration theory and practice: Part 1: Taxonomy of neurophysiological processes. *Sensory Integration Special Interest Section Quarterly*. 2000;23(1):1-4.
2. Mervis CB, John AE. Cognitive and behavioral characteristics of children with Williams syndrome: implications for intervention approaches. En: *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*. Wiley Online Library; 2010. p. 229-48.
3. Dunn W. Implementing neuroscience principles to support habilitation and recovery. *Occupational therapy: Enabling function and well-being*. 1997;182-233.
4. Lane SJ. Structure and function of the sensory systems. 2002.
5. Schaaf RC, Schoen SA, Smith Roley S, Lane SJ, Koomar J, May-Benson TA. A frame of reference for sensory integration. *Frames of reference for pediatric occupational therapy*. 2010;3:99-186.
6. Bar-Shalita T, Goldstand S, Hahn-Markowitz J, Parush S. Typical children's responsivity patterns of the tactile and vestibular systems. *American Journal of Occupational Therapy*. 2005;59(2):148-56.
7. Lane SJ. Sensory modulation. 2002.
8. Bar-Shalita T, Vatine J-J, Parush S. Sensory modulation disorder: A risk factor for participation in daily life activities. *Developmental Medicine & Child Neurology*. 2008;50(12):932-7.

9. Bundy AC, Shia S, Qi L, Miller LJ. How does sensory processing dysfunction affect play? *American Journal of Occupational Therapy*. 2007;61(2):201-8.
10. Dunn W. The impact of sensory processing abilities on the daily lives of young children and their families: A conceptual model. *Infants and young children*. 1997;9:23-35.
11. Engel-Yeger B, Ziv-On D. The relationship between sensory processing difficulties and leisure activity preference of children with different types of ADHD. *Research in developmental disabilities*. 2011;32(3):1154-62.
12. Nadon G, Feldman DE, Dunn W, Gisel E. Association of sensory processing and eating problems in children with autism spectrum disorders. *Autism research and treatment*. 2011;2011.
13. Reynolds S, Bendixen RM, Lawrence T, Lane SJ. A pilot study examining activity participation, sensory responsiveness, and competence in children with high functioning autism spectrum disorder. *Journal of autism and developmental disorders*. 2011;41(11):1496-506.
14. Schaaf RC, Miller LJ, Seawell D, O'Keefe S. Children with disturbances in sensory processing: A pilot study examining the role of the parasympathetic nervous system. *American Journal of Occupational Therapy*. 2003;57(4):442-9.
15. Case-Smith J, O'Brien JC. *Occupational therapy for children and adolescents-e-book*. Elsevier Health Sciences; 2014.

16. Dunn W. The sensations of everyday life: Empirical, theoretical, and pragmatic considerations. *American Journal of Occupational Therapy*. 2001;55(6):608-20.
17. Costa-López B, Ferrer-Cascales R, Ruiz-Robledillo N, Albaladejo-Blázquez N, Bary\la-Matejczuk M. Relationship between Sensory Processing and Quality of Life: A Systematic Review. *Journal of Clinical Medicine*. 2021;10(17):3961.
18. McIntosh DN, Miller LJ, Shyu V, Dunn W. The sensory profile: Examiner's manual. Psychological Corporation. 1999;
19. White BP, Mulligan S, Merrill K, Wright J. An examination of the relationships between motor and process skills and scores on the sensory profile. *American Journal of Occupational Therapy*. 2007;61(2):154-60.
20. Gándara-Gafo B, Santos-del Riego S, Viana-Moldes I, Muñoz J. Cultural Adaptation of the Adolescent/Adult Sensory Profile for Spain. *American Journal of Occupational Therapy*. 2019;73(6):7306205070p1-9.
21. Gándara-Gafo B, Santos-del Riego S, Muñoz J. Reference Values for the Adolescent/Adult Sensory Profile in Spain. *American Journal of Occupational Therapy*. 2019;73(5):7305205040p1-8.
22. Murray MM, Eardley AF, Edginton T, Oyekan R, Smyth E, Matusz PJ. Sensory dominance and multisensory integration as screening tools in aging. *Scientific reports*. 2018;8(1):1-11.

23. Brown C, Dunn W. Adult/adolescent sensory profile: User's manual. San Antonio, Texas: Psychological Corporation. 2002;
24. Dunn W. Sensory profile: User's manual. Psychological Corporation San Antonio, TX; 1999.
25. Kinnealey M, Oliver B. Adult sensory questionnaire. Unpublished raw data Temple University, college of allied health professionals, department of occupational therapy. 2002;3307.
26. Kinnealey M, Oliver B, Wilbarger P. A phenomenological study of sensory defensiveness in adults. *American Journal of Occupational Therapy*. 1995;49(5):444-51.
27. Kinnealey M, Koenig KP, Smith S. Relationships between sensory modulation and social supports and health-related quality of life. *American Journal of Occupational Therapy*. 2011;65(3):320-7.
28. Kaiser HF. The application of electronic computers to factor analysis. *Educational and psychological measurement*. 1960;20(1):141-51.
29. Schumacker RE, Lomax RG. A beginner's guide to structural equation modeling. psychology press; 2004.
30. Browne MW, Cudeck R. Alternative ways of assessing model fit. *Sociological methods & research*. 1992;21(2):230-58.
31. Tabachnick BG, Fidell LS. Using multivariate statistics. Allyn and Bacon. Needham Heights, MA. 2001;

32. Pereira JCR. Análise de dados qualitativos: estratégias metodológicas para as ciências da saúde humanas e sociais. Edusp; 1999.
33. Cattell RB. The scree test for the number of factors. *Multivariate behavioral research*. 1966;1(2):245-76.
34. Pfeiffer B, Kinnealey M. Treatment of sensory defensiveness in adults. *Occupational Therapy International*. 2003;10(3):175-84.
35. Johnson M-E, Irving R. Implications of sensory defensiveness in a college population. 2008;
36. Panagopoulos VN, Greene DJ, Campbell MC, Black KJ. Towards objectively quantifying sensory hypersensitivity: a pilot study of the “Ariana effect”. *PeerJ*. 2013;1:e121.
37. Ermer J, Dunn W. The Sensory Profile: A discriminant analysis of children with and without disabilities. *American Journal of Occupational Therapy*. 1998;52(4):283-90.
38. Rogers SJ, Hepburn S, Wehner E. Parent reports of sensory symptoms in toddlers with autism and those with other developmental disorders. *Journal of autism and developmental disorders*. 2003;33(6):631-42.
39. Ben-Sasson A, Hen L, Fluss R, Cermak SA, Engel-Yeger B, Gal E. A meta-analysis of sensory modulation symptoms in individuals with autism spectrum disorders. *Journal of autism and developmental disorders*. 2009;39(1):1-11.

40. Crane L, Goddard L, Pring L. Sensory processing in adults with autism spectrum disorders. *Autism*. 2009;13(3):215-28.
41. Horder J, Wilson CE, Mendez MA, Murphy DG. Autistic traits and abnormal sensory experiences in adults. *Journal of autism and developmental disorders*. 2014;44(6):1461-9.
42. Glod M, Riby DM, Honey E, Rodgers J. Sensory atypicalities in dyads of children with autism spectrum disorder (ASD) and their parents. *Autism Research*. 2017;10(3):531-8.
43. Bijlenga D, Tjon-Ka-Jie JYM, Schuijers F, Kooij JJS. Atypical sensory profiles as core features of adult ADHD, irrespective of autistic symptoms. *European Psychiatry*. 2017;43:51-7.
44. Sharfi K, Rosenblum S. Sensory modulation and sleep quality among adults with learning disabilities: A quasi-experimental case-control design study. *PloS one*. 2015;10(2):e0115518.
45. Rieke EF, Anderson D. Adolescent/Adult Sensory Profile and obsessive-compulsive disorder. *American Journal of Occupational Therapy*. 2009;63(2):138-45.
46. Engel-Yeger B, Muzio C, Rinosi G, Solano P, Geoffroy PA, Pompili M, et al. Extreme sensory processing patterns and their relation with clinical conditions among individuals with major affective disorders. *Psychiatry research*. 2016;236:112-8.

47. Beaudry-Bellefeuille I, Lane SJ. Cultural adaptation for Spain of the Spanish version of the Short Sensory Profile using cognitive interviews. *Austin Journal of Autism and Related Disabilities*. 2015;1:1004-10.
48. Gándara Gafo B. Adaptación cultural, validación y valores de referencia del "Adolescent/Adult Sensory Profile" en España. 2016;
49. Brown C, Tollefson N, Dunn W, Cromwell R, Filion D. The adult sensory profile: Measuring patterns of sensory processing. *American Journal of Occupational Therapy*. 2001;55(1):75-82.
50. Chung JCC. Measuring sensory processing patterns of older Chinese people: Psychometric validation of the adult sensory profile. *Aging and Mental Health*. 2006;10(6):648-55.
51. Almomani FM, Brown C, Dahab SA, Almomani M, Nadar M. Cross cultural adaptation of the adolescent/adult sensory profile: establishing linguistic equivalency and psychometric properties of the Arabic version. *Disability and rehabilitation*. 2014;36(9):765-70.
52. Kayihan H, Akel BS, Salar S, Huri M, Karahan S, Turker D, et al. Development of a Turkish version of the sensory profile: translation, cross-cultural adaptation, and psychometric validation. *Perceptual and motor skills*. 2015;120(3):971-86.
53. Al-Momani F, Alghadir AH, Al-Momani MO, Alharethy S, Al-Sharman A, Al-Dibii R, et al. Performance of the Arabic Population on the Adolescent-Adult Sensory Profile: An Observational Study. *Neuropsychiatric disease and treatment*. 2020;16:35.

54. Zaree M, Hassani Mehraban A, Lajevardi L, Saneii S, Pashazadeh Azari Z, Mohammadian Rasnani F. Translation, reliability and validity of Persian version of Adolescent/Adult Sensory Profile in dementia. *Applied Neuropsychology: Adult*. 2021;1-7.
55. Williams ZJ, Failla MD, Gotham KO, Woynaroski TG, Cascio C. Psychometric evaluation of the short sensory profile in youth with autism spectrum disorder. *Journal of autism and developmental disorders*. 2018;48(12):4231-49.
56. Chojnicka I, Pisula E. Adaptation and psychometric properties of the Polish version of the Short Sensory Profile 2. *Medicine*. 2019;98(44).
57. Nunnally JC. *Psychometric Theory* 2nd ed. Mcgraw hill book company; 1978.
58. Fayers PM, Machin D. *Quality of Life Assessment, Analysis and Interpretation* Chichester, England. John Wiley & Sons, Ltd; 2000.
59. Nunnally JC. *Psychometric theory* 3E. Tata McGraw-hill education; 1994.

VI. DISCUSIÓN

1. METODOLOGÍA COSMIN PARA REVISIONES SISTEMÁTICAS

A través de la totalidad de artículos incluidos en esta tesis empleamos la metodología COSMIN para revisiones sistemáticas, enumerando y evaluando las propiedades psicométricas de los instrumentos de medida para el estado nutricional (“Nutritional status measurement instrument: a systematic psychometric review”), el conocimiento dietético en población diabética (“The diabetes knowledge measurement instrument: a systematic psychometric review”) y la fatiga relacionada con el cáncer en cuidados paliativos (“Cancer-related fatigue measures in palliative care: a psychometric systematic review”).

Para estos 3 artículos y mediante la guía COSMIN hacemos uso de la metodología diseñada para revisiones sistemáticas sobre cuestionarios existentes, sobre los cuales se dispone de al menos alguna información sobre sus propiedades de medición, su uso con fines evaluativos y su selección para un propósito específico, a través de una serie de pasos: búsqueda y selección sistemática de artículos relevantes, valoración de la calidad de los estudios elegibles, evaluación de las propiedades de medición de los instrumentos incluidos, aspectos de interpretabilidad y factibilidad, formulación de conclusiones y recomendaciones, y su publicación (1).

Los primeros puntos incluyen una serie de procedimientos estándar que cumplen con las directrices existentes cuando se realizan revisiones sistemáticas, mientras que en los puntos posteriores se realiza una evaluación de las propiedades de medición (proceso más complejo), para finalizar con la selección del cuestionario.

Esa evaluación de las propiedades de medición contempla tres pasos:

La calidad metodológica de cada estudio sobre una propiedad de medición se evalúa utilizando la lista de verificación de riesgo de sesgo de COSMIN (2), la calidad de los resultados para cada estudio sobre una propiedad de medición se califica en función de los criterios actualizados para las buenas propiedades de

medición de Terwee y Prinsen (3,4) y, en último lugar, la evidencia se resume y la calidad de esta se califica mediante el método GRADE modificado (3–5).

1.1 COSMIN

Algunos autores de revisiones sistemáticas sobre propiedades de medición han evaluado la calidad metodológica en distintos estudios (6–9), utilizando diferentes métodos para ello: Haywood et al. (6) consideró el número, tipo de pruebas y estudios realizados; Alla et al. (7) empleó criterios diseñados para la evaluación de la calidad de los ensayos; Marinus et al. (8) usó la idoneidad de los análisis y el tamaño de la muestra, y Wind et al. (9) tuvo en cuenta el objetivo, la población, el método de evaluación, los análisis y la presentación de los resultados estadísticos.

Igualmente, debemos saber que existen una gran cantidad de instrumentos de medición de resultados (muchos de ellos miden un mismo constructo y están diseñados para la misma población de estudio), y aún se están desarrollando otros nuevos. El proceso de selección de estas herramientas de evaluación para propósitos específicos es complejo. Las opciones implican consideraciones conceptuales, como definir el constructo y la población; aspectos prácticos, como la carga para los pacientes y evaluadores, además de los costos, y aspectos de calidad evaluados por 9 propiedades de medición diferentes agrupadas en los dominios de fiabilidad, validez y sensibilidad al cambio (10).

Sobre la base de las puntuaciones obtenidas con estos instrumentos de medida se toman decisiones sobre la aplicación de posteriores pruebas diagnósticas y tratamientos, por lo que deben ser fiables y válidos. De lo contrario, existe un riesgo grave de resultados imprecisos que podrían conducir a sesgos en las conclusiones de los estudios, produciendo un derroche de recursos y una falta de ética en relación con los pacientes participantes, padeciendo estos las cargas y riesgos del estudio (11).

Organizaciones como Food and Drug Administration (FDA) y European Medicines Agency (EMA) requieren que estos instrumentos de medición estén bien validados para su propósito (12,13). Marshall et al. (14) demostró la necesidad de instrumentos de medición fiables y validos relacionados con la

salud y expuso, en una revisión sobre esquizofrenia, una asociación entre el uso de escalas no publicadas en ensayos clínicos y la obtención de resultados significativos. Concluyó que los ensayos tenían más probabilidad de informar que un tratamiento era superior al control cuando se usaba una escala no validada para la evaluación de los participantes de un estudio. Y es que el hecho de utilizar herramientas sin ser validadas previamente podría influir en los resultados obtenidos y mostrar una mayor efectividad en un tratamiento debido a una evaluación incorrecta.

Antes de que un instrumento diseñado para ser utilizado en ciencias de la salud pueda emplearse en la investigación o la práctica clínica, se deben evaluar y considerar la totalidad de las propiedades de medición posibles para el instrumento en cuestión: validez de contenido, validez estructural, consistencia interna, fiabilidad, error de medición, prueba de hipótesis, validez transcultural/invariancia de medición, validez de criterio y sensibilidad al cambio; o en su defecto aquellas agrupadas en los 3 dominios: validez, fiabilidad y sensibilidad al cambio (10). Los estudios que evalúen estas propiedades de medición deben ser de alta calidad metodológica para garantizar conclusiones apropiadas sobre las propiedades psicométricas de esta nueva herramienta.

La iniciativa COSMIN basándose en el consenso alcanzado por un amplio panel de expertos a nivel internacional (estudio Delphi internacional) (15–17), pretende mejorar la selección de estos instrumentos de medición de resultados mediante el desarrollo de pautas metódicas para evaluar la calidad metodológica de los estudios sobre las propiedades de medición, tanto en la investigación como en la práctica clínica (1,18).

Este manual se desarrolló de acuerdo con las guías existentes para revisiones: el manual Cochrane para revisiones sistemáticas de intervenciones (19) y para revisiones de precisión de pruebas diagnósticas (20); la Declaración PRISMA (21); la lista de verificación STARD (22); las normas del Instituto de Medicina para revisiones sistemáticas de la investigación de la eficacia comparativa (23); los criterios actualizados para las buenas propiedades de medición de Terwee y Prinsen (3,4) y la metodología para formular y calificar recomendaciones en la práctica clínica GRADE (3–5).

El desarrollo de dicho manual con todas estas guías validadas internacionalmente garantiza que este cumple con todos los requisitos que asegurarían la realización de revisiones sistemáticas de instrumentos de valoración con una alta calidad metodológica, así como una correcta selección de los instrumentos para evaluar el constructo en cuestión.

Se fundamenta en un protocolo para realizar revisiones sistemáticas sobre instrumentos de medición, basado en una guía de 10 puntos (<https://www.cosmin.nl/>) (24). En dicho protocolo se describe el método de síntesis de la mejor evidencia en el que se puede combinar el número de estudios, su calidad y la consistencia de los resultados para determinar la solidez de la evidencia para cada propiedad de medición. Se obtienen pruebas concluyentes de una fiabilidad adecuada cuando se encuentran resultados positivos consistentes (ICC o Kappa > 0.70) en al menos 2 estudios de buena calidad o un estudio de excelente calidad, procedimiento similar al empleado en la metodología GRADE. Los valores de corte desarrollados previamente (como ICC o Kappa > 0.70) se utilizan para determinar si un instrumento de medición de resultados tiene buenas propiedades de medición (25).

A pesar de contar con una adecuada fiabilidad, un alfa de Cronbach alto no garantiza que se esté midiendo el constructo de interés o que no falte ningún concepto importante (26,27), y una fiabilidad o capacidad de respuesta test-retest alta no implica que todos los ítems sean relevantes y que no se hayan perdido conceptos fundamentales. Uno puede medir el constructo incompleto o incorrecto de manera muy fiable y un cambio real en el constructo de interés puede sobreestimarse o subestimarse debido a conceptos irrelevantes o faltantes. De igual manera, los pacientes pueden frustrarse cuando se hacen preguntas que les parecen irrelevantes o no son trascendentes, lo que puede dar lugar a respuestas sesgadas o tasas de respuesta bajas (28,29).

Es importante tener en cuenta que existen diversos factores que pueden diferenciar una “muy buena o buena” fiabilidad de una fiabilidad “dudosa o inadecuada”, como la estabilidad de los pacientes con respecto al constructo que se medirá entre las diferentes administraciones (3), un apropiado intervalo de tiempo y similitud entre las pruebas (3,28), la selección del método estadístico adecuado para el estudio de la fiabilidad test-retest (ICC, Pearson o Spearman)

(28,30), un uso apropiado de Kappa para evaluar la concordancia o reproducibilidad de instrumentos de medida cuyo resultado es categórico (Kappa de Cohen (para puntajes dicotómicos o nominales) o Kappa ponderado (para escalas ordinales)) (28,31,32), y/o administraciones no realizadas de forma independiente (3).

Antes de medir la fiabilidad, hemos de tener en cuenta otra propiedad psicométrica más importante, la validez de contenido. La guía de la FDA sobre los resultados informados por los pacientes recomienda establecer la validez de contenido antes de evaluar otras propiedades de medición (13). COSMIN hace lo propio y también sugiere considerar primero dicha propiedad al evaluar y comparar las propiedades de medición de los instrumentos de evaluación en una revisión sistemática (33).

Messick enfatizó la importancia de la relevancia del contenido y la cobertura de las pruebas educativas para determinar lo que los estudiantes han aprendido en un curso. Cada elemento de la prueba debe relacionarse con uno de los objetivos del curso, y cada parte de este debe estar representada por una o más preguntas (34). Los mismos principios se aplican al contenido de una herramienta de evaluación: todos sus elementos deben ser relevantes para el constructo de interés (dentro de una población específica y un contexto de uso), dicho instrumento debe ser integral con respecto a las preocupaciones del paciente (28,35–38) y debe ser entendido por este según lo previsto.

En un reciente estudio Delphi a nivel internacional sobre la selección de instrumentos de evaluación para un conjunto mínimo de resultados acordados (COS) (39), se llegó al consenso de que al menos la validez de contenido y la estructura interna (validez estructural y consistencia interna, y/o IRT/ Rasch) tenían que ser adecuadas para recomendar una herramienta de medición para un conjunto de resultados básicos que debían medirse e informarse para todos los ensayos clínicos en relación con una enfermedad específica o población de estudio. Estas afirmaciones aprobadas por consenso merecen ser analizadas, ya que una escala con unos excelentes valores para estas propiedades psicométricas podría no ser recomendable, como si, por ejemplo, obtuviera valores bajos de fiabilidad inter-observador, lo cual nos haría sospechar que dicha escala puede

obtener diferentes resultados dependiendo de quién sea el profesional que la evalúe.

La validez de contenido de los instrumentos de valoración existentes se puede evaluar preguntando a los pacientes y profesionales sobre la relevancia, la exhaustividad y la comprensibilidad de los elementos, las opciones de respuesta y las instrucciones (28,40). Sin embargo, los métodos utilizados varían ampliamente y muchos estudios solo abordan la comprensibilidad, sin prestar atención a la relevancia y la exhaustividad (41), por lo que se necesitan nuevas directrices para evaluar la calidad metodológica de los estudios de validez de contenido. En COSMIN sucede lo mismo, ya que el recuadro que contiene los estándares sobre validez de contenido necesita ser actualizado básicamente por tres razones: no contiene estándares para evaluar la calidad del desarrollo de un instrumento de evaluación, no se presta atención a la comprensibilidad de la herramienta y los estándares únicamente hacen referencia a si se hicieron ciertas cosas, pero no a cómo se hicieron (16).

De igual modo, se necesitan criterios de lo que constituye una buena validez de contenido para brindar recomendaciones transparentes y basadas en la evidencia para la selección de una herramienta de evaluación en revisiones sistemáticas para determinar si un instrumento de evaluación es lo suficientemente bueno para medir los resultados, para la aprobación regulatoria de un nuevo medicamento o para la inclusión en un registro de herramientas de evaluación (16).

Esta validez de contenido se puede calificar utilizando los criterios de revisión de Scientific Advisory Committee of the Medical Outcomes Trust (SAC-MOT) (42,43), la herramienta Evaluating Measures of Patient Reported Outcomes (EMPRO) (44), los criterios de buenas propiedades publicados por Terwee et al. (4) o los estándares mínimos recomendados por International Society for Quality of Life Research (ISOQOL) (45). Sin embargo, estos solo se definen de manera general y no incluyen criterios específicos para calificar la relevancia, la exhaustividad y la comprensibilidad de manera estandarizada, ya que aún no existen métodos para combinar la evidencia del estudio de desarrollo de un instrumento de evaluación y estudios adicionales de validez de contenido en una revisión sistemática.

Es importante destacar que además de la metodología COSMIN (3,4,24) para evaluar la calidad de estos instrumentos de evaluación que miden un concepto específico en una población determinada, existen otras guías como los estándares de American Psychological Association (APA) (46,47), FDA (13,48), Agency for Healthcare Research and Quality (HRQL) (49,50), SAC-MOT (42,43), la herramienta EMPRO (basada en SAC-MOS) (44), los criterios de Terwee y Prinsen (1,3), una lista de verificación publicada en 2016 por Francis et al. (51), la extensión CONSORT-PRO (52), los estándares mínimos recomendados de ISOQOL (45), Guidelines for Reporting Reliability and Agreement Studies (GRRAS) (53), Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) (54), Quality Appraisal of Diagnostic Reliability (QAREL) (55), etc.

La diferencia de estas guías con respecto a COSMIN radica en que la mayoría son mucho más breves, sus pautas a seguir no describen ni explican todos los pasos necesarios para determinar la calidad de este tipo de instrumentos de evaluación de manera sistemática, en sus directrices no se establece que la conclusión deba basarse en una consideración sistemática de toda la evidencia existente, combinan estándares para la calidad de los estudios con criterios para calificar los resultados de las propiedades de medición en un solo estándar y estos estándares no están lo suficientemente detallados para proporcionar calificaciones transparentes y sistemáticas del riesgo de sesgo de los estudios sobre dichas propiedades de medición.

Otras apreciaciones a nivel individual entre estas guías y COSMIN podrían residir en que estas no tratan los estudios que investigan las propiedades psicométricas (plantea los resultados informados por los pacientes utilizados en los ensayos clínicos) (52), no abordan el informe de estudios que prueban la propiedad de medida de un instrumento de evaluación (45), no evalúan todas las propiedades psicométricas (43,46,47,53), no describen cómo deben combinarse los resultados de diferentes estudios con distinta calidad para sacar una conclusión general sobre un instrumento de evaluación (1,3,51), no se centran en instrumentos de medición de resultados (evalúan la calidad de los estudios sobre la precisión de pruebas diagnósticas) (54,55), en algunas directrices no se incluye en el estándar de consistencia interna el requisito de valorar la unidimensionalidad (análisis factorial) (44,51), calculan la validez de constructo

sin requerir ninguna información sobre la calidad de los instrumentos de evaluación ya existentes (51) y algunas propiedades de medición no se mencionan específicamente, entre ellas fiabilidad y error de medición (42–44,51).

1.2 TERWEE

De igual manera, COSMIN determina que es imprescindible conocer la calidad de los resultados para cada estudio sobre una propiedad de medición. En estas revisiones sistemáticas, tradicionalmente, se comparan las propiedades de contenido y medición de los cuestionarios disponibles. Se necesitan razonamientos para determinar la calidad metodológica de los estudios sobre el desarrollo y evaluación de cuestionarios en relación con el estado de salud y criterios de buenas propiedades de medición para legitimar cuál es el mejor.

Aun así, cabe destacar que una escala con mejores propiedades en cuanto a la calidad de sus resultados no debe ser considerada mejor que otra similar que mida el mismo constructo, puesto que en la revisiones sistemáticas presentes en esta tesis doctoral han sido añadidas escalas con excelentes valores en cuanto a la calidad de sus resultados y, en cambio, estas no han sido recomendadas para la evaluación del constructo en cuestión.

Como criterios de buenas propiedades de medición, probablemente los más conocidos y completos sean los de SAC-MOT (43). Este comité asesor científico definió 8 atributos de las propiedades de los instrumentos que justifican su consideración en la evaluación: modelo conceptual y de medición, validez, fiabilidad, sensibilidad al cambio, interpretabilidad, demandas y carga administrativa, formas alternativas y cultura y adaptaciones del idioma (traducciones). Dentro de cada uno de estos atributos se definieron criterios específicos por los cuales los instrumentos deben ser revisados.

Existen otros criterios similares y estos han sido definidos por algunos autores como McDowell y Jenkinson (56), Andresen (57) y Bombardier y Tugwell (58). Sin embargo, lo que a menudo le falta a estos criterios son razonamientos explícitos sobre lo que constituyen buenas propiedades de medición. Por ejemplo, sobre cuántas hipótesis deben confirmarse para justificar que un cuestionario tiene una buena validez, para aquello que constituye un buen acuerdo (error de

medición aceptable), una buena capacidad de respuesta, una buena interpretabilidad, un tamaño de muestra adecuado para los estudios que evalúan las propiedades de medición, etc.

Terwee asume el desafío de analizar y perfeccionar más aún los criterios de calidad disponibles para los estudios sobre el desarrollo y la evaluación de cuestionarios sobre el estado de salud, incluidos criterios explícitos para las actuales propiedades de medición: validez de contenido, validez estructural, consistencia interna, fiabilidad, error de medición, prueba de hipótesis, validez transcultural/invariancia de medición, validez de criterio y sensibilidad al cambio (1,3,39). Para ello, considera dos aspectos a tener en cuenta: en primer lugar, que a través de estos criterios los cuestionarios de baja calidad pueden recibir calificaciones positivas para algunas propiedades de medición y en segundo lugar, que aquellos cuestionarios con el mayor número de valoraciones positivas no son necesariamente los mejores, ya que estas calificaciones dependen de la disponibilidad de información y la calidad de los informes sobre la evaluación de las propiedades de medición (59).

Terwee probó inicialmente sus criterios originales (anteriores a los actuales) (59) para dos revisiones sistemáticas que compararon las propiedades de medición de los cuestionarios para la discapacidad del hombro (60) y para el funcionamiento visual (61) y concluyó que, tras aplicarlos, varias propiedades de medición a menudo no se evaluaban, analizaban o transmitían adecuadamente y la información importante para la validez de contenido solía ser muy limitada. Algunas de las razones podrían fundamentarse en que estos criterios se basan principalmente en opiniones, puesto que no existe evidencia empírica en este campo que respalde criterios de calidad explícitos, sino que se trata de reglas generales útiles que en algunas ocasiones hacen que otros investigadores prefieran tomar sus propias decisiones.

Estos criterios de buenas propiedades de medición no se resumen en una puntuación de calidad general como suele hacerse en las revisiones sistemáticas de ECAs (62), puesto que una puntuación de calidad general asume que todas las propiedades de medición son igualmente importantes, lo que probablemente no sea cierto, y es que para Terwee la validez de contenido es la propiedad psicométrica más significativa (al igual que ocurriera en la lista de verificación de

riesgo de sesgo de COSMIN) porque solo si la validez de contenido de un cuestionario es adecuada, se considerará el uso de este y será útil la evaluación del resto de sus propiedades.

A la hora de confeccionar unos criterios sólidos se deben tener en cuenta una serie de consideraciones:

Un estudio de validación mal informado conducirá a calificaciones bajas para cuestionarios sobre el estado de salud que no son necesariamente deficientes en diseño o rendimiento, indicativo de la necesidad de mejorar la calidad de los estudios sobre el desarrollo y la evaluación de estos.

Es de vital importancia saber si el cuestionario fue diseñado con fines discriminatorios o evaluativos (algo que a menudo no se describe), si los conceptos que pretende medir están bien definidos y si la selección de elementos se describe de manera correcta (63), porque dicho proceso facilitará el juicio sobre la aplicabilidad del cuestionario en una población específica.

Es importante asegurarse de que el número de escalas se corresponde con el número de dimensiones identificadas en el análisis factorial, ya que a veces descubrimos que varios cuestionarios afirmaban cubrir más de una dimensión, pero consistían en una sola escala o viceversa (60).

Hay que comprobar si las hipótesis para la evaluación de la validez de constructo han sido detalladas (muchos autores no lo hacen) o si estas no son lo suficientemente específicas o informativas. Una hipótesis debe evaluar tan directamente como sea posible unas afirmaciones determinadas formuladas para una teoría (64) y, por lo tanto, debe incluir una justificación de la dirección esperada y la magnitud de la correlación (sin probar hipótesis específicas, el riesgo de sesgo es alto). Muchos de los autores de estudios de evaluación de instrumentos determinan que su instrumento es válido, mientras que el recuento objetivo del número de hipótesis que se confirmaron indica lo contrario (60,61).

Por último, la sensibilidad al cambio a menudo está mal definida y no está bien evaluada, algunos autores utilizan innumerables índices de sensibilidad al cambio (diferencias conceptuales y algebraicas entre estos múltiples índices) (65,66) y deducen que su instrumento responde, a pesar de que todos los índices apuntan en la misma dirección (67-69).

1.3 GRADE

Además de la calidad metodológica y la calidad de los resultados, en la metodología COSMIN se considera fundamental evaluar la calidad de la evidencia. Para su evaluación, el primer paso consiste en la automatización de la búsqueda de la evidencia para la elaboración de la mejor certeza de la que disponemos, teniendo siempre en cuenta que, a medida que nos aproximemos a ella, debemos de considerar la presencia de características culturales, poblacionales, económicas, ambientales, tecnológicas, etc. (validez externa). El siguiente paso radica en proporcionarle un significado a esa evidencia, para lo cual será necesario seleccionar la clasificación más acorde a nuestra necesidad y aquella que haga posible descartar entre un nivel de certeza u otro para efectuar las recomendaciones más apropiadas al entorno poblacional y asistencial.

Existen variadas proposiciones basadas en diferentes métodos para clasificar la evidencia: desde la originaria y más sencilla, Sackett (70); hasta otras más novedosas, complejas y extensas como HRQL (49,50), Canadian Asthma Consensus Group (CACG) (71,72), la Agencia de Evaluación de Tecnología Médica (AATM) (73), la del Servicio Andaluz de Salud (SAS) (74), Strength of Recommendation Taxonomy (SORT) (75), North American Nursing Diagnosis Association (NANDA-I) (76); con capacidad para discernir en diferentes situaciones (National Health and Medical Research Council (NHMRC) (77), Centre for Evidence-Based Medicine (CEBM) (78), National Institute for Health and Clinical Excellence (NICE) (79), Practicing Chiropractors' Committee on Radiology Protocols (PCCRP) (80)); clasificaciones más definidas, que señalen a una situación en concreto (Scottish Intercollegiate Guidelines Network (SIGN) (81), Canadian Task Force on Preventive Health Care (CTFPHC) (82), U.S. Preventive Services Task Force (USPSTF) (83), American Diabetes Association (ADA) (84), American College of Cardiology Foundation / American Heart Association (ACCF/AHA) (85)), y otras como GRADE (5) que ha sido incluida para valorar la calidad de la evidencia disponible, llevar a cabo recomendaciones y crear guías de práctica clínica fundamentadas en la evidencia por aproximadamente 90 instituciones de gran reputación a nivel nacional e internacional (86).

COSMIN recomienda el uso exclusivamente de GRADE en revisiones sistemáticas y guías porque tiene claras ventajas con respecto al resto de estos métodos. Primero, aumenta la reproducibilidad en comparación con enfoques menos sistemáticos. Segundo, proporciona un marco para todo el proceso de realización de una revisión sistemática o desarrollo de una guía, el cual requiere que todos los pasos se realicen de manera sistemática y transparente (la idoneidad de cada paso del proceso depende de la realización sistemática y transparente del paso anterior). En tercer lugar, las evaluaciones tienen implicaciones directas tanto para la práctica como para la investigación y las evaluaciones para la calidad de la evidencia resaltan dónde la base de la evidencia es adecuada o dónde se necesita más o mejor investigación. Por último, el empleo generalizado de este método, la amplia orientación disponible y el empleo de un lenguaje sencillo en la certeza de las evaluaciones y recomendaciones de la evidencia hacen que este sea reconocible y fácil de usar e interpretar (87).

COSMIN utiliza un enfoque GRADE modificado (adopta solo 4 de los 5 factores de este método) para calificar la calidad de la evidencia en revisiones sistemáticas sobre instrumentos de evaluación. Se emplea para degradar la calidad de la evidencia cuando hay dudas sobre la fiabilidad de los resultados, basándose en estos 4 factores: riesgo de sesgo (calidad metodológica de los estudios), inconsistencia (incongruencia inexplicable de los resultados entre los estudios), imprecisión (tamaño total de la muestra de los estudios disponibles) y ausencia de evidencia directa (evidencia de poblaciones diferentes a la población de interés en la revisión) (1).

La calidad de la evidencia se califica para cada propiedad de medición y para cada instrumento de evaluación por separado. El punto de partida es siempre la suposición de que el resultado conjunto o general es de alta calidad, para posteriormente degradar dicha calidad en uno o dos niveles por factor (a moderada, baja o muy baja) cuando existe riesgo de sesgo, inconsistencia, imprecisión o ausencia de evidencia directa (es útil considerar todos los factores uno por uno usando el orden consecutivo).

En la revisión sistemática sobre FRC en CP, uno de los instrumentos evaluados mediante el método GRADE modificado fue la escala SAS. En primer lugar, se evaluó el riesgo de sesgo para el único estudio donde fue incluida.

Teniendo en cuenta que esta escala tenía una calidad metodológica dudosa y una calidad de los resultados suficiente, se redujo el nivel de evidencia por riesgo de sesgo de alto a moderado (-1), pero este hecho podría conducir a conclusiones precipitadas, ya que hablamos de un solo estudio para esta escala en este tipo de población.

En segundo lugar, después del riesgo de sesgo, se valoró la inconsistencia (siempre debemos considerar una mayor degradación debido a otros factores). Al haber obtenido una calificación suficiente para la calidad de los resultados, no fue necesaria una rebaja por inconsistencia, de lo contrario se debería haber considerado la degradación. En el supuesto de que esta escala hubiera obtenido una calificación insuficiente para la calidad de los resultados, habríamos tenido que rebajar otro punto dentro del factor inconsistencia (-1), puntualizando que podemos añadir lo anteriormente comentado en cuanto a la falta de estudios que hagan fehaciente esta afirmación.

El siguiente paso hubiera sido tener en cuenta el tamaño de la muestra. En esta escala es superior a 100 (concretamente 572 pacientes), por lo que no hubiera sido necesario rebajar otro punto (-1), ya que GRADE recomienda bajar un nivel cuando el tamaño total de la muestra de los estudios agrupados o resumidos es inferior a 100 y dos niveles cuando el tamaño total de la muestra está por debajo de 50 (3).

Lo que ocurre en estos casos es que GRADE otorga una calidad alta a una muestra mayor de 100. Pero esta muestra no garantiza una calidad óptima, ya que para algunas poblaciones, por ejemplo, pacientes con ictus en los que existen millones de personas con esta afección, una muestra de más de 100 participantes puede ser insuficiente, mientras que si se evalúa una enfermedad rara, reclutar a más de 100 sujetos sería casi imposible, debido principalmente a que la incidencia de dicha patología es ínfima.

Para el último de los factores, ausencia de evidencia directa, la certeza podría ser degradada un punto más, ya que en el caso de esta escala en concreto, en su población de estudio recoge un pequeño tanto por ciento de pacientes sanos (un 5%), y los criterios de inclusión están diseñados para pacientes con cáncer y en tratamiento en cuidados paliativos. Además, hemos de tener en cuenta lo puntualizado con anterioridad para los dos primeros factores (riesgo de sesgo e

inconsistencia), hablamos de un solo estudio y no hay mucho que comparar en cuanto a características poblacionales se refiere.

2. METODOLOGÍA COSMIN PARA VALIDACIÓN DE ESCALAS

Aparte de la lista de verificación de riesgo de sesgo de COSMIN para revisiones sistemáticas, la guía también contiene unos estándares de validación para cuestionarios de evaluación en ciencias de la salud, con el objetivo de generalizar el proceso de validación de dichos cuestionarios siguiendo una misma metodología y evaluando cada una de las propiedades psicométricas recogidas en esta (24). En dicha tesis y siguiendo estos estándares se llevó a cabo la validación de la versión española del cuestionario “Validation of the Adolescent-Adults Sensory Profile (AASP) for Spain”, a través de la ejecución de una serie de propiedades psicométricas (validez estructural, consistencia interna, fiabilidad, error de medición y validez de criterio convergente), mostrando la metodología empleada y los resultados obtenidos.

Al igual que ocurre en las revisiones sistemáticas, para la validación de una escala COSMIN considera la validez de contenido como la propiedad psicométrica más importante. En nuestro trabajo de investigación no fue posible medir esa propiedad, ni tampoco la validez transcultural, puesto que el cuestionario en cuestión ya estaba traducido y validado transculturalmente a población española (88,89).

2.1 VALIDEZ ESTRUCTURAL

Algunos investigadores exponen que los análisis estadísticos sobre las características de las escalas y los ítems también forman parte de la evaluación de la validez de contenido, y otros investigadores no. Un grupo de trabajo de la iniciativa Patient-Reported Outcomes Measurement Information System (PROMIS) consideró que la asignación de los ítems formaba parte de la evaluación de la validez de contenido (90). Por el contrario, el grupo de trabajo International Society for Pharmacoeconomics and Outcomes Research

(ISPOR) no los incluyó como parte de la evaluación de esta propiedad psicométrica (35–37,91,92). Dentro de la taxonomía COSMIN, los análisis estadísticos como la escala de elementos y el análisis factorial se consideran parte de la evaluación de la consistencia interna, y la validez estructural y la calidad metodológica de dichos estudios se evalúa mediante casillas separadas (18).

Estamos de acuerdo con Magasi et al. (90) en que probar la estructura interna de una herramienta de evaluación es esencial, y que puede apuntar a elementos que no miden un mismo constructo. De tal manera, COSMIN recomienda evaluar la validez de contenido primero y de forma independiente, para posteriormente determinar la validez estructural y finalmente la consistencia interna.

La validez estructural suele evaluarse mediante análisis factorial o análisis IRT/Rasch. Para valorar la validez estructural se efectúa un análisis factorial en todos los ítems de un instrumento de evaluación para calcular el número (hipótesis) de subescalas de este y la agrupación de ítems dentro de las subescalas (estudios de validez estructural). Para valorar la unidimensionalidad por subescala se realizan análisis de factores múltiples en los ítems de cada subescala por separado, para evaluar si cada subescala por sí sola mide un solo constructo (estudios de unidimensionalidad). Estos estudios son suficientes para la interpretación de los análisis de consistencia interna y para los de IRT/Rasch, pero no proporcionan evidencia de validez estructural como parte de la validez de constructo (1).

El panel Delphi decidió que la evaluación de la validez estructural se podía realizar mediante EFA o CFA. Sin embargo, se prefiere el CFA al EFA porque el CFA comprueba si los datos se ajustan a una estructura factorial hipotética *a priori*, mientras que el EFA se puede utilizar cuando no existen hipótesis claras sobre las dimensiones subyacentes (no es una herramienta sólida en la prueba de hipótesis) (93).

En la lista de verificación de COSMIN no se discuten los requisitos específicos para los análisis factoriales, como la elección del EFA (análisis de componentes principales o análisis factorial común), la elección y justificación del método de rotación (rotación ortogonal u oblicua) o la decisión sobre el número

de factores relevantes; requisitos específicos que si se argumentan en otros estudios, como el de Floyd y Widaman (93) o el de De Vet et al. (94).

En nuestro trabajo de investigación se ejecutó un EFA con extracción de máxima verosimilitud y rotación varimax para la estructura interna del nuevo cuestionario, a través del cual se extrajeron 6 factores. Posteriormente, se realizó un CFA para determinar la validez estructural del mismo. Para ello, se calcularon las cargas factoriales de las variables y se empleó la estimación de máxima verosimilitud para considerar el ajuste del modelo (se incluyeron chi cuadrado (χ^2), RMSEA y CFI). El resultado final consistió en una reducción del número de ítems de la versión original (60 ítems) con respecto a la versión española del nuevo cuestionario (42 ítems). Todo este proceso estadístico para el cálculo del análisis factorial demuestra la importancia de seguir una serie de estándares como los recogidos en la guía COSMIN.

2.2 CONSISTENCIA INTERNA

La consistencia interna es una propiedad de medición importante para los cuestionarios que pretenden medir un solo concepto subyacente (constructo) mediante el uso de múltiples elementos. Por el contrario, para cuestionarios en los que los ítems son simplemente diferentes aspectos de un fenómeno clínico complejo y no tienen porqué estar correlacionados, como en el caso de la Escala de Apgar (96), la consistencia interna no es relevante (97,98).

Cuando esta propiedad de medición es relevante, se debe aplicar el análisis de componentes principales o el análisis factorial para determinar si los ítems forman exclusivamente una escala general (dimensión) o más de una (93,99). En caso de que no exista una hipótesis previa sobre la dimensionalidad de un cuestionario, se puede aplicar el análisis exploratorio de componentes principales o el análisis factorial. Pero si existe una hipótesis clara con respecto a la estructura factorial se debe utilizar el CFA (94,100). En nuestro estudio y siguiendo la metodología COSMIN para la escala analizada, al contener esta varias dimensiones se calculó la consistencia interna para cada una de ellas, así como para la escala en general.

Después de determinar el número de escalas o subescalas se debe calcular el alfa de Cronbach (medida adecuada de consistencia interna) para cada escala o subescala por separado. Un alfa de Cronbach bajo indica falta de correlación entre los ítems de una escala, mientras uno muy alto indica altas correlaciones entre los ítems de la escala (redundancia de uno o más ítems). Nunnally y Bernstein propusieron un criterio de 0.70-0.90 como medida de buena consistencia interna (101). Para nuestra investigación, los valores resultantes fueron óptimos: el α de Cronbach para cada uno de los 4 cuadrantes fue 0.70, 0.69, 0.73 y 0.73, y los valores para el cuestionario completo oscilaron entre $r = 0.816 - 0.831$, resultando para el total de los ítems un $\alpha = 0.83$.

2.3 FIABILIDAD

Delphi llegó a un consenso sobre los métodos estadísticos preferidos para evaluar la fiabilidad de las medidas de resultado que tienen puntuaciones continuas, ordinales y dicotómicas o nominales, respectivamente, y cómo deben ser calificados. Los métodos estadísticos elegidos fueron ICC y Kappa de Cohen ponderado (17).

El ICC es el parámetro de fiabilidad más adecuado y utilizado para medidas continuas. En la mayoría de los casos, muchos autores no describen qué ICC han empleado, si un ICC para consistencia (ICC consistency) o un ICC para el grado de concordancia (ICC agreement) (28,102). Para McGraw y Wong (102) el método elegido es el ICC agreement (modelo de efectos aleatorios de dos vías o el ICC (A,1)), debido a que las diferencias sistemáticas se consideran parte del error de medición.

Por su parte, el coeficiente Kappa de Cohen ponderado se utiliza para medidas ordinales (cuando se usan ponderaciones cuadráticas, el coeficiente Kappa de Cohen ponderado es idéntico al ICC agreement) (28). El coeficiente de correlación de Pearson no se emplea, ya que se considera inadecuado porque no se tienen en cuenta las diferencias sistemáticas; al igual que el porcentaje absoluto de concordancia, el cual no se ajusta por la correspondencia atribuible al azar (28).

Siguiendo las directrices COSMIN para el artículo recogido en la presente tesis doctoral, se optó por llevar a cabo un análisis estadístico mostrando el ICC

(3,1) con el objetivo de exponer el componente estadístico más adecuado a dicho fin. A pesar de que el coeficiente de correlación de Pearson también fue utilizado, este se empleó para mostrar diferentes propiedades psicométricas, no para determinar la fiabilidad. El resultado final mostró un ICC con un intervalo de una semana entre ambas pruebas para cada uno de los 4 cuadrantes de 0.90, 0.90, 0.90 y 0.83, y los valores para cada uno de los ítems variaron entre $r = 0.40-0.92$

2.4 ERROR DE MEDICIÓN

El panel Delphi eligió el SEM como el método idóneo para medir el error de medición. Al preguntar sobre otros métodos apropiados, solo un pequeño porcentaje (20%) estuvo de acuerdo con SDC y con LoA (28%). A pesar de los bajos porcentajes de acuerdo alcanzados sobre la aceptación de SDC y LoA como métodos válidos, el comité directivo decidió posteriormente que ambos procedimientos también deberían considerarse apropiados para medir el error de medición y se incluyeron en la lista de verificación (15).

Delphi expuso que para puntuaciones continuas, el SEM, SDC o LoA eran los métodos estadísticos elegidos, mientras que para puntuaciones ordinales, dicotómicas o nominales era el porcentaje de concordancia específica (positivo y negativo). Conjuntamente, el comité consideró el coeficiente de variación (CV) como el método estadístico preferido para evaluar el error de medición para escalas con puntuaciones de proporción o porcentaje y propuso que este fuese aceptado de igual modo como un método estadístico más para calcular el error de medición (17).

COSMIN estableció como “muy bueno” un error de medición en el que se calculaba el SEM, SDC o LoA para puntuaciones continuas, o el porcentaje de concordancia (positivo y negativo) para puntuaciones dicotómicas, nominales u ordinales (95). Y al mismo tiempo, consideró como “inadecuado” un SEM calculado a partir de un alfa de Cronbach o usando el SD de otra población, ya que a veces se emplea el alfa de Cronbach en lugar del ICC para calcular el error de medición (40). Igualmente, el alfa de Cronbach a veces se usa de manera incorrecta porque se calcula para una escala que no es unidimensional o se basa en un modelo formativo, contribuyendo a que este no se pueda interpretar (17).

Para nuestro estudio y siguiendo la metodología COSMIN, se evaluó el SEM de cada una de las variables incluidas en los ítems de la AASP-42, y este osciló entre 0.026 - 0.051.

2.5 VALIDEZ DE CRITERIO

El panel Delphi llegó al consenso de que no existe el “gold standard” para los instrumentos de evaluación relacionados con la salud, y discutió si la propiedad de medición validez de criterio debía mencionarse en la lista de verificación de COSMIN. Decidió que la única excepción para un “gold standard” se da cuando se compara un instrumento abreviado con la versión larga original (18) y siguiendo estos criterios comparó la escala resultante con menor número de ítems, con la escala original en español.

2.6 PRUEBA DE HIPÓTESIS

La prueba de hipótesis establece si la dirección y la magnitud de una correlación o diferencia es similar a lo que se podría esperar en función de los constructos que se están midiendo. Cuantas más hipótesis se prueben sobre si los datos corresponden a hipótesis formuladas *a priori*, más evidencia se reunirá para la validez del constructo.

El panel Delphi acordó que las hipótesis específicas que han de ser probadas, deben formularse *a priori* sobre las diferencias medias esperadas entre grupos conocidos o las correlaciones esperadas entre las puntuaciones del instrumento y otras variables. Igualmente, la dirección esperada (positiva o negativa) y la magnitud (absoluta o relativa) de las correlaciones o diferencias deben incluirse para dichas hipótesis (64,103,104). En nuestro caso, nuestra hipótesis *a priori* consistía en que la nueva escala de procesamiento sensorial con un menor número de ítems obtendría excelentes valores de correlación con respecto a la escala ya traducida al español, hipótesis que quedó confirmada tras el análisis de los resultados.

Las hipótesis también pueden hacer referencia a la magnitud relativa de las correlaciones o a las diferencias en las puntuaciones entre grupos. Al evaluar dichas diferencias, es menos trascendente cuando estas son estadísticamente

significativas (dependen del tamaño de la muestra) o cuando tienen la magnitud esperada.

El panel Delphi recomendó que se eviten los valores de “p” en las hipótesis, porque no es relevante examinar si las correlaciones o las diferencias difieren estadísticamente de 0 (105). El tamaño de la diferencia es más importante que las diferencias significativas entre los grupos, debido a que depende del número de sujetos para cada grupo. Es preferible que las pruebas de hipótesis formales se basen en la magnitud esperada de las correlaciones y diferencias, en lugar de los valores de “p”. Cuando se formulan hipótesis sobre las relaciones esperadas con otros instrumentos, estos instrumentos de comparación deben describirse adecuadamente, y sus propiedades de medición deberían haberse evaluado en la misma versión de idioma y población de pacientes en la que se llevó a cabo el estudio.

Para la lista de verificación de COSMIN, la propiedad psicométrica prueba de hipótesis va a ser clasificada dentro de un recuadro u otro, en función de si el instrumento de medida de resultados se comparó con otros instrumentos de evaluación similares (validez convergente) o para un mismo instrumento entre subgrupos (validez discriminativa) (95). Debido a que nuestra escala no ha sido evaluada en diferentes subgrupos (solo en personas sanas) se optó por que fuera comparada con un instrumento similar, siendo en este caso la misma escala con un mayor número de ítems (escala original).

2.7 SENSIBILIDAD AL CAMBIO

La sensibilidad al cambio es una medida de validez longitudinal. En analogía con la validez de constructo, la validez longitudinal debe evaluarse probando hipótesis predefinidas sobre las correlaciones esperadas entre los cambios en las medidas o las diferencias esperadas en los cambios entre grupos "conocidos" (106). Esto muestra la idoneidad de un cuestionario para medir cambios si realmente han ocurrido y su capacidad para distinguir cambios clínicamente importantes en relación con errores de medición.

El panel Delphi quería incluir la sensibilidad al cambio como una propiedad de medición separada. Durante la discusión, estuvo de acuerdo en que la única

diferencia entre la validez transversal (de constructo y de criterio) y la sensibilidad al cambio era que la validez hacía referencia a la validez de una puntuación única y la sensibilidad al cambio a la validez de una puntuación de cambio (10), por lo que decidió que los estándares de esta propiedad de medida debían ser análogos a los estándares de validez de criterio y de constructo.

Las medidas adecuadas para evaluar la sensibilidad al cambio son las mismas que para la prueba de hipótesis y la validez de criterio, con la única excepción de que las hipótesis deben centrarse en la puntuación de cambio de un instrumento. De Boer et al. evaluó la sensibilidad al cambio del cuestionario Low Vision Quality of Life (LVQOL) y Vision Related Quality of Life Core Measure (VCM1) mediante pruebas de hipótesis pre-especificadas sobre las relaciones de los cambios en los cuestionarios en relación con variaciones en otras medidas en pacientes con pérdida de visión irreversible (107). Formularon una hipótesis que consistía en que “la correlación del cambio en LVQOL/VCM1 con la modificación en el Visual Functioning questionnaire (VF-14) era mayor que la correlación con la escala de calificación global, el cambio en la agudeza visual y el cambio en el termómetro Euroqol”. Después de calcular las correlaciones entre las puntuaciones de cambio en los diferentes instrumentos de medición, concluyeron si las correlaciones eran las esperadas.

Existen una serie de parámetros propuestos en la literatura para evaluar la capacidad de respuesta que el panel Delphi consideró inapropiados. Llegó a un consenso por el cual el uso del tamaño del efecto (108) y otros métodos relacionados, como la respuesta media estandarizada (109), el coeficiente de sensibilidad al cambio S de Norman (110) y la estadística de eficiencia relativa (111) eran medidas inadecuadas para medir la sensibilidad al cambio; al igual que la prueba t pareada (medida de cambio significativo, en lugar de válido, y método dependiente del tamaño de la muestra) (105) o el estadístico de sensibilidad al cambio de Guyatt (tiene en cuenta el MIC, y el panel Delphi determinó que este no hace referencia a la validez de la puntuación de cambio) (112).

Para el estudio de investigación recogido en la tesis doctoral se evaluó la sensibilidad al cambio siguiendo esta misma metodología. En el caso de la escala Sensory Profile, esta propiedad psicométrica cobra vital importancia porque existen patrones para los que los nuevos cuadrantes indican diferencias de

escasos 3 o 4 puntos, hecho que le da aún más significación a esta propiedad, ya que podría hacer que una misma persona evaluada dos veces con esta misma escala pueda estar al mismo tiempo en un rango u otro de puntuación.

3. LIMITACIONES

Algunos de los problemas que plantea la lista de verificación de riesgo de sesgo de COSMIN estarían relacionados con su complejidad (116 elementos en 10 categorías), la cual podría limitar su utilidad para un revisor sistemático, investigador o profesional sanitario que no tenga experiencia en propiedades de medición; su uso manifiesto, dado que este se reduce a instrumentos de evaluación diseñados para aplicaciones en las que se pretenden medir los cambios a lo largo del tiempo; su falta de aplicabilidad para instrumentos discriminatorios, herramientas que se utilizan con fines predictivos o para las medidas relacionadas con la atención de la salud (aquellas empleadas para evaluar la satisfacción con la atención o la adherencia); los habituales efectos suelo, puesto que estudios considerados como "muy buenos" al principio, posteriormente son calificados con una puntuación menor e incluso como "inadecuados" por un solo ítem (consecuencia del método empleado para determinar la puntuación), y el uso de los mismos ítems para medir las propiedades psicométricas de fiabilidad intra, inter-observador y test-retest.

Por ello, una metodología simplificada que incorpore las características críticas destacadas en COSMIN y una adecuada literatura, sería útil para permitir que los revisores sistemáticos, investigadores y profesionales sanitarios evalúen las características de desarrollo y la utilidad de una amplia variedad de instrumentos de evaluación. Además, su utilidad se vería mejorada por la inclusión de aspectos prácticos sobre las medidas de estas herramientas de evaluación que no se abordan de manera consistente en otros criterios (13).

Muchos de los revisores que afirman aplicar su lista de verificación no son expertos en el desarrollo o evaluación de medidas, estos pueden ser especialistas en aspectos clínicos de la enfermedad estudiada, pero este no es el objetivo que se

pretende, puesto que la mayoría de los instrumentos de evaluación han sido desarrollados por personas con capacitación clínica que tienen un interés particular en las herramientas de evaluación que miden el estado de salud, y este constructo es muy diferente al de juzgar si los pacientes se beneficiaran de las intervenciones.

Algunos autores afirman la existencia de diversos sesgos identificados al evaluar las revisiones basadas en COSMIN, consecuencia de la aplicación de la lista de verificación por parte de revisores no cualificados (113–115). Quizá COSMIN debería especificar qué revisiones cumplen con sus criterios de calidad y cuáles son los motivos para llevarlas a cabo, dado que con demasiada frecuencia los revisores sistemáticos realizan una investigación para demostrar la necesidad de una nueva medida, algo sin sentido, ya que si encuentran que una medida actual es de alta calidad no hay necesidad de una nueva valoración. No obstante, hay varias razones por las que se necesitan estas revisiones, como relacionar el constructo ya cubierto por el instrumento de evaluación con los resultados específicos requeridos por un estudio, saber si existe suficiente calidad y capacidad de respuesta para detectar cambios en el constructo medido o conocer el grado de aceptabilidad por parte de los encuestados. Esta diversidad de necesidades puede ser otra limitación, debido a que podría confundir a los lectores.

Otro de los inconvenientes sería que la lista de verificación contiene algunas pautas sobre cómo juzgar la calidad de los instrumentos de evaluación ya revisados. Si bien los profesionales sanitarios y otros investigadores pueden decir que un tercero imparcial ha llegado a la conclusión de que un instrumento es bueno, no hay garantía de que ese tercero sea imparcial. Esto es particularmente problemático cuando los revisores no logran distinguir los diferentes tipos de instrumentos de evaluación y esto repercute en que algunas herramientas de evaluación reconocidas por su desempeño deficiente suelen calificarse como buenas, pero es posible que otros revisores las califiquen como malas (por ejemplo, en distintos estudios sobre dos herramientas de evaluación: Short Form-36 Health Survey (SF-36) y sus variantes, y las diferentes versiones de EuroQol-5D (EQ-5D), ambos cuestionarios obtuvieron buenas calificaciones (116–119) y malas al mismo tiempo (120–123)), por lo que los resultados de las revisiones

basadas en COSMIN, generalmente, no están respaldados por la evidencia presentada por los revisores.

A este respecto, Szekeres (124) hizo algunos comentarios positivos sobre la lista de verificación de COSMIN, sin embargo, concluyó que “la validez de contenido de la lista de verificación de COSMIN solo ha sido establecida por los desarrolladores y debe ser evaluada por expertos imparciales”, e informó que la fiabilidad entre evaluadores de esta lista de verificación es baja, situándose el porcentaje de acuerdo para las calificaciones por debajo del 80% para un tercio de las recomendaciones (125).

En relación con los criterios de buenas propiedades publicados por Terwee y Prinsen, algunas de las limitaciones presentes están fundamentadas en que, en estos criterios, no se describe cómo deben combinarse los resultados de diferentes estudios con distinta calidad para sacar una conclusión general sobre un instrumento de evaluación; al aplicar sus criterios algunas propiedades de medición a menudo no se evalúan, analizan o transmiten de una forma adecuada (en los criterios actuales no se detalla que los valores alfa de Cronbach > 0.95 para la consistencia interna han de considerarse negativos, puesto que son indicativos de redundancia entre los ítems) y la información importante para la validez de contenido suele estar muy limitada. Todo ello, teniendo en cuenta que estos criterios se basan principalmente en opiniones, dado que no existe evidencia empírica en este campo que respalde unos criterios de calidad explícitos.

Aparte de las limitaciones encontradas en la lista de verificación de COSMIN y los criterios de buenas propiedades publicados por Terwee y Prinsen, GRADE también describe una serie de inconvenientes a la hora de aplicar su metodología: este se ha desarrollado para abordar preguntas sobre estrategias, intervenciones o políticas de gestión alternativas, no para preguntas sobre el riesgo o el pronóstico; el intento de aplicación de dicho método a un conjunto mal definido de recomendaciones, conocidas como “recomendaciones de buenas prácticas” (estas recomendaciones pueden no ser útiles y cuando lo son, puede que no sea un ejercicio ventajoso calificar la calidad de la evidencia o calificar la fuerza de las recomendaciones); la falta de conocimiento por parte de revisores y desarrolladores de guías para comprender dónde encaja dicho método en el proceso general (la preparación de una guía implica varias pautas antes y después

de los pasos a los que se aplica el método GRADE) (19,126); la ausencia de experiencia de este en relación con la evaluación de cuestiones de salud pública, pruebas diagnósticas y sistemas de salud (quienes apliquen GRADE a preguntas sobre estos temas se enfrentarán a algunos desafíos especiales) (127,128) y, para finalizar, añadir también que este método supondrá una decepción para aquellos que esperan un marco que elimine los desacuerdos al interpretar la evidencia y al decidir cuál es el mejor entre los cursos de acción alternativos (aunque este método expone juicios sobre la calidad de la evidencia y la fuerza de las recomendaciones de una manera más sistemática y transparente, no elimina la necesidad de emitir esos juicios).

Para concluir y en relación con nuestro trabajo de investigación, sería importante comentar 3 aspectos sobre el desarrollo de la metodología COSMIN en la validación de instrumentos de evaluación. En primer lugar, podemos asegurar que, para la aplicación de dicho método al proceso de validación de una escala, son necesarios amplios conocimientos estadísticos. En segundo lugar, para nuestro artículo se ha usado la escala en español ya traducida y adaptada culturalmente a la lengua española (88,89), pero se desconoce si el proceso de traducción se ha realizado siguiendo una metodología reproducible y mediante el procedimiento indicado por COSMIN. En último lugar, el procesamiento sensorial al ser una variable “nueva”, no existe aún un “gold standard” definido que evalúe dicha escala, por lo que la escala validada para nuestro artículo de investigación podría ser la “gold standard”, hecho que aún se desconoce.

4. IMPLICACIONES PARA LA PRÁCTICA CLÍNICA, PARA LA INVESTIGACIÓN Y RECOMENDACIONES PARA FUTURAS LÍNEAS DE INVESTIGACIÓN

En ausencia de evidencia empírica y en respuesta al problema que supone seleccionar la herramienta de evaluación más adecuada, el comité directivo de COSMIN desarrolló una metodología para realizar revisiones sistemáticas con el objetivo de medir un constructo de interés en una población de estudio específica y seleccionar el instrumento de evaluación más apropiado. Este tipo de revisiones sistemáticas pueden proporcionar una descripción general completa de las propiedades de medición de estos cuestionarios, determinar el riesgo de sesgo de los resultados de los estudios individuales incluidos en dicha revisión y respaldar las recomendaciones basadas en la evidencia para la selección del instrumento más adecuado para un propósito determinado.

Estos instrumentos de evaluación van a ayudar a identificar aspectos relevantes referentes a problemas de salud presentes en el individuo, como aquellos relacionados con el funcionamiento físico, emocional, social, de rol funcional, cognitivo, de percepción de la salud general y el bienestar, y en relación con síntomas específicos para cada enfermedad.

Por ello, es necesario tener un conocimiento exhaustivo sobre la metodología COSMIN y su aplicación para la realización de revisiones sistemáticas sobre un constructo de interés en una población de estudio específica, con el fin de posibilitar la selección del instrumento más apropiado y facilitar la validación de este tipo de herramientas, ya que con el instrumento idóneo se puede dar respuesta a un problema de salud aportando a los profesionales del ámbito sanitario un recurso ineludible de gran aplicación clínica que ayude a favorecer el proceso de valoración, diagnóstico, planificación de estrategias, tipos de intervención y su evaluación, con los beneficios que ello reporta en cuanto a salud, recursos humanos y materiales. En este caso han sido presentados los instrumentos idóneos para evaluar ingesta dietética y conocimiento dietético en diabetes, así como fatiga en cáncer, por lo que los resultados mostrados en la presente revisión sistemática serían de vital

importancia para la práctica clínica a la hora de evaluar dichas variables con la herramienta de valoración correcta.

COSMIN también presenta una serie de patrones para cuestionarios de evaluación en ciencias de la salud con el objetivo de estandarizar el proceso de validación de dichos cuestionarios, siguiendo una misma metodología y evaluando cada una de las propiedades psicométricas recogidas en su guía, pautas seguidas en nuestro artículo de investigación para la validación del cuestionario Sensory Profile en población española.

Para futuras investigaciones sería interesante especificar y consensuar las pautas metodológicas estandarizadas a desarrollar durante la validación de un instrumento de evaluación en relación con los métodos ya existentes, mediante un comité de expertos o una comisión científica basándose en la evidencia existente sobre propiedades de medición y su posterior difusión a la comunidad científica con el objetivo de contrarrestar el exceso de información tan variada que existe sobre el tema. Adicionalmente, se aconseja que las revisiones ofrezcan información específica sobre la metodología empleada en el proceso, con el fin de ofrecer la mayor transparencia posible.

Sería sugerente la creación de programas de aprendizaje para los profesionales interesados en manejar la guía COSMIN, tanto para fines investigadores como clínicos. Asimismo, con el fin de hacer más fácil la aplicación de dicha guía, sería de gran interés plantear un manual que brindara un resumen de las pautas a seguir para su aplicación, tanto en revisiones sistemáticas como en el proceso de validación de herramientas de evaluación.

VII. CONCLUSIONES

VII. CONCLUSIONES

- En la revisión sistemática llevada a cabo sobre herramientas de evaluación dirigidas a evaluar el estado nutricional en pacientes con diabetes mellitus tipo 1 y/o 2 utilizando la metodología COSMIN, el cuestionario Food Frequency Questionnaire fue el instrumento de evaluación incluido en mayor volumen de estudios, con más características psicométricas evaluadas y con una mejor puntuación en cuanto a la calidad de la evidencia, lo que lo convierte en el cuestionario recomendado para evaluar la ingesta dietética en personas con diabetes.

- En la revisión sistemática llevada a cabo sobre herramientas de evaluación dirigidas a medir el conocimiento dietético en pacientes con diabetes mellitus tipo 1 y/o 2 utilizando la metodología COSMIN, los cuestionarios Michigan Diabetes Knowledge Test, Diabetes Knowledge Questionnaire y Personal Diabetes Questionnaire fueron los instrumentos aconsejados para evaluar el conocimiento dietético en personas con diabetes mellitus tipo 1 y 2. Estos 3 cuestionarios fueron los que se incluyeron en un mayor volumen de estudios, en los que más características psicométricas se evaluaron y los que presentaron mejor puntuación en cuanto a calidad de la evidencia.

- En la revisión sistemática llevada a cabo sobre herramientas de evaluación dirigidas a evaluar la fatiga relacionada con el cáncer en cuidados paliativos utilizando la metodología COSMIN, los cuestionarios Edmonton Symptom Assessment System, Problems and Needs in Palliative Care Questionnaire y European Organization for Research and Treatment of Cancer Quality of Life 15 items Questionnaire for Palliative Care fueron los cuestionarios con mayor presencia en cuanto a número de estudios y propiedades psicométricas evaluadas, lo que hace que estos 3 cuestionarios sean los más ampliamente validados y los más recomendados para este tipo de población. Para esta evaluación también se incluyó Palliative Care Quality of Life Instrument, ya que a pesar de haber sido usado en un solo estudio, presentó las mejores

puntuaciones en cuanto a calidad metodológica y calidad de resultados para las propiedades psicométricas más significativas.

- Para la validación del cuestionario Adolescent/Adult Sensory Profile en población española, aplicando los estándares para propiedades psicométricas recogidos en la guía COSMIN, los resultados obtenidos mostraron una estructura factorial para una versión de 42 ítems (Adolescent/Adult Sensory Profile- 42) recogidos en 4 subescalas, un α de Cronbach de 0.83 para la totalidad de la escala y un ICC para cada una de las 4 subescalas de 0.90, 0.90, 0.90 y 0.83. Se compararon las puntuaciones de la versión original de Adolescent/Adult Sensory Profile- 60 con los ítems de la nueva versión adaptada al español de Adolescent/Adult Sensory Profile- 42, y los niveles de correlación encontrados fueron muy buenos ($r = 0.88-0.97$; $p < 0.01$). Por lo que, tras analizar dichas propiedades de medida, podemos concluir que Adolescent/Adult Sensory Profile- 42 es una herramienta válida y fiable con óptimas propiedades psicométricas para evaluar el procesamiento sensorial en dicha población.

VIII. BIBLIOGRAFÍA

VIII. BIBLIOGRAFÍA

1. Food US, Administration D. Guidance for Industry. US Department of Health and Human Services. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. 2018.
2. Health UD of, Services H. Guidance for industry-Patient-reported outcome measures: Use in medical product development to support labeling claims. <http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm>. 2009;
3. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of clinical epidemiology*. 2010;63(7):737-45.
4. COSMIN - Improving the selection of outcome measurement instruments [Internet]. COSMIN. [citado 21 de diciembre de 2021]. Disponible en: <https://www.cosmin.nl/>
5. Prinsen CA, Mokkink LB, Bouter LM, Alonso J, Patrick DL, De Vet HC, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research*. 2018;27(5):1147-57.
6. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies

-
- on measurement properties of health status measurement instruments: an international Delphi study. *Quality of life research*. 2010;19(4):539-49.
7. Mokkink LB, Prinsen C, Patrick DL, Alonso J, Bouter LM, De Vet HC, et al. COSMIN methodology for systematic reviews of patient-reported outcome measures (PROMs). User manual. 2018;78(1).
 8. Terwee CB, Prinsen CA, Chiarotto A, De Vet H, Bouter LM, Alonso J, et al. COSMIN methodology for assessing the content validity of PROMs—user manual. Amsterdam: VU University Medical Center. 2018;
 9. Association AP, Association AER. National Council on Measurements Used in Education.(1954). Technical recommendations for psychological tests and diagnostic techniques. *Psychological Bulletin*. 51(2 Pt 2):1-38.
 10. Eignor DR. The standards for educational and psychological testing. 2013;
 11. Patrick DL, Burke LB, Powers JH, Scott JA, Rock EP, Dawisha S, et al. Patient-reported outcomes to support medical product labeling claims: FDA perspective. *Value in Health*. 2007;10:S125-37.
 12. Feeny DH, Eckstrom E, Whitlock EP, Perdue LA. A primer for systematic reviewers on the measurement of functional status and health-related quality of life in older adults. 2013;
 13. Lohr KN. Assessing health status and quality-of-life instruments: attributes and review criteria. *Quality of life Research*. 2002;11(3):193-205.

14. Valderas JM, Ferrer M, Mendivil J, Garin O, Rajmil L, Herdman M, et al. Development of EMPRO: a tool for the standardized assessment of patient-reported outcome measures. *Value in Health*. 2008;11(4):700-8.
15. Francis DO, McPheeters ML, Noud M, Penson DF, Feurer ID. Checklist to operationalize measurement characteristics of patient-reported outcome measures. *Systematic reviews*. 2016;5(1):1-11.
16. Badia X, Carné X. La evaluación de la calidad de vida en el contexto del ensayo clínico. *Med Clin (Barc)*. 1998;110(14):550-6.
17. Arribas M. Diseño y validación de cuestionarios. *Matronas profesión*. 2004;5(17):23-9.
18. Carvajal A, Centeno C, Watson R, Martínez M, Rubiales AS. How is an instrument for measuring health to be validated? En: *Anales del sistema sanitario de Navarra*. 2011. p. 63-72.
19. Alcaraz FG, Espín AA, Martínez AH, Alarcón MM. Diseño de Cuestionarios para la recogida de información: metodología y limitaciones. *Revista clinica de medicina de familia*. 2006;1(5):232-6.
20. Argimon Pallás JM, Jiménez Villa J. *Métodos de investigación: clínica y epidemiológica*. 2000.
21. de Yébenes Prous MJG, Salvanés FR, Ortells LC. Validation of questionnaires. *Reumatología Clínica (English Edition)*. 2009;5(4):171-7.

22. Alconero-Camarero AR, Gualdrón-Romero A, Sarabia-Cobo CM, Martínez-Arce A. Clinical simulation as a learning tool in undergraduate nursing: Validation of a questionnaire. *Nurse Education Today*. 2016;39:128-34.
23. Choi BC, Pak AW. Peer reviewed: a catalog of biases in questionnaires. *Preventing chronic disease*. 2005;2(1).
24. Grau G. Metodología para la validación de cuestionarios. *Medifam*. 1995;5(6):351-9.
25. Gómez Benito J, María DH. La validez en los tests, escalas y cuestionarios. 2015;
26. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR task force for translation and cultural adaptation. *Value in health*. 2005;8(2):94-104.
27. Polit DF, Hungler BP, Palacios Martínez R, Feher de la Torre G. Investigación científica en ciencias de la salud: principios y métodos. 2000.
28. Tovar J. Psicometría: tests psicométricos, confiabilidad y validez. *Psicología: Tópicos de actualidad*. 2007;8.
29. Ara MJN, Saboya PR. Métodos, diseños y técnicas de investigación psicológica. Universidad Nacional de Educación a Distancia; 2006.
30. Cobo CMS, Camarero ARA. Claves para el diseño y validación de cuestionarios en Ciencias de la Salud. *Enfermería en cardiología: revista*

- científica e informativa de la Asociación Española de Enfermería en Cardiología. 2019;(77):69-73.
31. Román CJVDH, Rodríguez-Marín J. Adaptación de un cuestionario para la medida de la representación de la enfermedad. *Psicothema*. 1997;237-45.
 32. Abad FJ. *Medición en ciencias sociales y de la salud*. Síntesis; 2011.
 33. Cronbach LJ. Coefficient alpha and the internal structure of tests. *psychometrika*. 1951;16(3):297-334.
 34. Nunnally JC. *Psychometric Theory*: 2d Ed. McGraw-Hill; 1978.
 35. Streiner DL, Norman GR, Cairney J. *Health measurement scales: a practical guide to their development and use*. Oxford University Press, USA; 2015.
 36. Crocker L, Algina J. *Introduction to classical and modern test theory*. ERIC; 1986.
 37. Muniz J, Elosua P, Hambleton RK. International Test Commission Guidelines for test translation and adaptation. *Psicothema*. 2013;25(2):151-7.
 38. Tornimbeni S, Pérez E, Olaz F, de Kohan NC, Fernández A, Cupani M. *Introducción a la psicometría*. Paidós Buenos Aires; 2008.
 39. Meneses J, Barrios M, Bonillo A, Cosculluela A, Lozano LM, Turbany J, et al. *Aproximación histórica y conceptos básicos de la psicometría*. Psicometría Barcelona: Editorial UOC. 2013;25-73.
 40. Feinstein AR. T. Duckett Jones Memorial Lecture. The Jones criteria and the challenges of clinimetrics. *Circulation*. 1982;66(1):1-5.

41. Fava GA, Tomba E, Sonino N. Clinimetrics: the science of clinical measurements. *International journal of clinical practice*. 2012;66(1):11-5.
42. Hofman J. Clinimetría y la actividad médica. *Revista Argentina de Reumatología*. 2008;2.
43. Mokkink LB, Prinsen CA, Patrick DL, Alonso J, Bouter LM, de Vet HC, et al. COSMIN Study Design checklist for Patient-reported outcome measurement instruments. Amsterdam, The Netherlands. 2019;
44. García Valdés M, Suárez Marín M. El método Delphi para la consulta a expertos en la investigación científica. *Revista Cubana de Salud Pública*. 2013;39(2):253-67.
45. Figueroa GA, Montilla MAC, Melo RM. Método DELPHI: aplicaciones y posibilidades en la gestión prospectiva de la investigación y desarrollo. *Revista Venezolana de Análisis de Coyuntura*. 2012;18(1):41-52.
46. Mokkink LB, Boers M, van der Vleuten CPM, Bouter LM, Alonso J, Patrick DL, et al. COSMIN Risk of Bias tool to assess the quality of studies on reliability or measurement error of outcome measurement instruments: a Delphi study. *BMC medical research methodology*. 2020;20(1):1-13.
47. Gagnier JJ, Lai J, Mokkink LB, Terwee CB. COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures. *Quality of Life Research*. 2021;1-22.
48. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. The COSMIN checklist for evaluating the methodological quality of studies

- on measurement properties: a clarification of its content. *BMC medical research methodology*. 2010;10(1):1-8.
49. Brito-Brito PR, Martín-García Á, Oter-Quintana C, Paloma-Castro O, Romero-Sánchez JM. Development and Content Validation of a NOC-Based Instrument for Measuring Dietary Knowledge in Patients with Diabetes: CoNOCidiet-Diabetes. *International journal of nursing knowledge*. 2019;
50. Gil JAP, Moscoso SC, Rodríguez RM. Validez de constructo: el uso de análisis factorial exploratorio-confirmatorio para obtener evidencias de validez. *Psicothema*. 2000;12(Su2):442-6.
51. Batista-Foguet JM, Coenders G, Alonso J. Confirmatory factor analysis. Its role on the validation of health related questionnaires. *Medicina clinica*. 2004;122:21-7.
52. Kaiser HF. The application of electronic computers to factor analysis. *Educational and psychological measurement*. 1960;20(1):141-51.
53. Cattell RB. The scree test for the number of factors. *Multivariate behavioral research*. 1966;1(2):245-76.
54. Tabachnick BG, Fidell LS. *Using multivariate statistics*. Allyn and Bacon. Needham Heights, MA. 2001;
55. Schumacker RE, Lomax RG. *A beginner's guide to structural equation modeling*. psychology press; 2004.
56. Fan X, Sivo SA. Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research*. 2007;42(3):509-29.

-
57. Browne MW, Cudeck R. Alternative ways of assessing model fit. *Sociological methods & research*. 1992;21(2):230-58.
 58. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*. 1999;6(1):1-55.
 59. Carretero-Dios H, Pérez C. Normas para el desarrollo y revisión de estudios instrumentales. *International Journal of clinical and health psychology*. 2005;5(3):521-51.
 60. Oviedo HC, Arias AC. Aproximación al uso del coeficiente alfa de Cronbach. *Revista colombiana de psiquiatría*. 2005;34(4):572-80.
 61. Campo-Arias A, Oviedo HC. Psychometric properties of a scale: internal consistency. *Revista de salud pública*. 2008;10(5):831-9.
 62. Pallás JMA, Villa JJ. *Métodos de investigación clínica y epidemiológica*. Elsevier; 2019.
 63. Tangarife JL, Arias JAC. Construcción y validación de escalas de medición en salud: revisión de propiedades psicométricas. *Archivos de medicina*. 2015;11(3):1.
 64. Soler Cárdenas SF. Coeficientes de confiabilidad de instrumentos escritos en el marco de la teoría clásica de los tests. *Educ med super*. 2008;
 65. da Silva FC, Gonçalves E, Arancibia BA, Bento GG, Castro TL, Hernandez SS, et al. Estimators of internal consistency in health research: the use of the

- alpha coefficient. *Revista peruana de medicina experimental y salud publica*. 2015;32(1):129-38.
66. Bland JM, Altman DG. Statistics notes: Cronbach's alpha. *Bmj*. 1997;314(7080):572.
67. Cárdenas SFS, Pons LS. Usos del coeficiente alfa de Cronbach en el análisis de instrumentos escritos. *Revista Médica Electrónica*. 2012;34(1):1-6.
68. Henson RK. Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and evaluation in counseling and development*. 2001;34(3):177-89.
69. Krus DJ, Helmstadter GC. The problem of negative reliabilities. *Educational and Psychological Measurement*. 1993;53(3):643-50.
70. Campo-Arias A. Usos del coeficiente de alfa de Cronbach. *Biomedica*. 2006;26(4):585-8.
71. Streiner DL. Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of personality assessment*. 2003;80(1):99-103.
72. Prieto L, Lamarca R, Casado A. Assessment of the reliability of clinical findings: the intraclass correlation coefficient. *Medicina clinica*. 1998;110(4):142-5.
73. Müller R, Büttner P. A critical discussion of intraclass correlation coefficients. *Statistics in medicine*. 1994;13(23-24):2465-76.

-
74. Ramada-Rodilla JM, Serra-Pujadas C, Delclós-Clanchet GL. Cross-cultural adaptation and health questionnaires validation: revision and methodological recommendations. *Salud publica de Mexico*. 2013;55(1):57-66.
75. Ortiz MS, Fernández-Pera M. Modelo de Ecuaciones Estructurales: Una guía para ciencias médicas y ciencias de la salud. *Terapia psicológica*. 2018;36(1):51-7.
76. Jackson DL, Gillaspay Jr JA, Purc-Stephenson R. Reporting practices in confirmatory factor analysis: an overview and some recommendations. *Psychological methods*. 2009;14(1):6.
77. Schmitt TA. Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment*. 2011;29(4):304-21.
78. Bou JC, Satorra A. A multigroup structural equation approach: A demonstration by testing variation of firm profitability across EU samples. *Organizational Research Methods*. 2010;13(4):738-66.
79. Byrne BM. Testing for multigroup equivalence of a measuring instrument: A walk through the process. *Psicothema*. 2008;872-82.
80. Bolin JH. Introduction to mediation, moderation, and conditional process analysis: a regression-based approach. *JSTOR*; 2014.
81. Saris WE, Satorra A, Van der Veld WM. Testing structural equation models or detection of misspecifications? *Structural Equation Modeling*. 2009;16(4):561-82.

82. Hu L, Bentler PM, Kano Y. Can test statistics in covariance structure analysis be trusted? *Psychological bulletin*. 1992;112(2):351.
83. Argibay JC. Técnicas psicométricas. Cuestiones de validez y confiabilidad. 2006;
84. AM AM, Muñoz NS. Some methodological issues about measurements in health. *Revista médica de Chile*. 2008;136(1):125-30.
85. DeVon HA, Block ME, Moyle-Wright P, Ernst DM, Hayden SJ, Lazzara DJ, et al. A psychometric toolbox for testing validity and reliability. *Journal of Nursing scholarship*. 2007;39(2):155-64.
86. Lamprea JA, Gómez-Restrepo C. Validez en la evaluación de escalas. *Revista Colombiana de psiquiatría*. 2007;36(2):340-8.
87. Streiner DL, Norman GR, Cairney J. *Health measurement scales: a practical guide to their development and use*. Oxford University Press, USA; 2015.
88. Sánchez Pedraza R, Gómez Restrepo C. Conceptos básicos sobre validación de escalas. *Rev colomb psiquiatr*. 1998;121-30.
89. Streiner DL. 22 A Checklist for Evaluating the Usefulness of Rating Scales. *A Guide for the Statistically Perplexed*. 2020;267-88.
90. Strauss ME, Smith GT. Construct validity: Advances in theory and methodology. *Annual review of clinical psychology*. 2009;5:1-25.
91. Sánchez R, Echeverry J. Validating scales used for measuring factors in medicine. *Revista de salud pública*. 2004;6(3):302-18.

92. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *Journal of chronic diseases*. 1987;40(2):171-8.
93. Maneesriwongul W, Dixon JK. Instrument translation process: a methods review. *Journal of advanced nursing*. 2004;48(2):175-86.
94. Mokkink LB, De Vet HC, Prinsen CA, Patrick DL, Alonso J, Bouter LM, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Quality of Life Research*. 2018;27(5):1171-9.
95. Griffiths C, Armstrong-James L, White P, Rumsey N, Pleat J, Harcourt D. A systematic review of patient reported outcome measures (PROMs) used in child and adolescent burn research. *Burns*. 2015;41(2):212-24.
96. Mokkink LB, Prinsen CA, Bouter LM, de Vet HC, Terwee CB. The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and how to select an outcome measurement instrument. *Brazilian journal of physical therapy*. 2016;20:105-13.
97. Terwee CB, Prinsen CAC, Garotti MR, Suman A, De Vet HCW, Mokkink LB. The quality of systematic reviews of health-related outcome measurement instruments. *Quality of Life Research*. 2016;25(4):767-79.
98. Higgins JP. *Cochrane handbook for systematic reviews of interventions*. Version 5.1. 0 [updated March 2011]. The Cochrane Collaboration. www.cochrane-handbook.org. 2011;

99. Bossuyt P, Davenport C, Deeks J, Hyde C, Leeflang M, Scholten R. Cochrane handbook for systematic reviews of diagnostic test accuracy. The Cochrane Collaboration. 2008;
100. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Int J Surg.* 2010;8(5):336-41.
101. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Annals of internal medicine.* 2003;138(1):W1-12.
102. Eden J, Levit L, Berg A, Morton S. Institute of Medicine; Board on Health Care Services; Committee on Standards for Systematic Reviews of Comparative Effectiveness Research. *Finding what works in health care: Standards for systematic reviews.* Retrieved February 27, 2017. 2011.
103. GRADE handbook [Internet]. [citado 16 de diciembre de 2021]. Disponible en: <https://gdt.gradepro.org/app/handbook/handbook.html>
104. Prinsen CA, Vohra S, Rose MR, Boers M, Tugwell P, Clarke M, et al. How to select outcome measurement instruments for outcomes included in a “Core Outcome Set”—a practical guideline. *Trials.* 2016;17(1):1-10.
105. Fayers PM, Hand DJ, Bjordal K, Groenvold M. Causal indicators in quality of life research. *Quality of life research.* 1997;6(5):393-406.
106. De Vet HC, Terwee CB, Mokkink LB, Knol DL. *Measurement in medicine: a practical guide.* Cambridge university press; 2011.

107. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of clinical epidemiology*. 2007;60(1):34-42.
108. De Vet HC, Adèr HJ, Terwee CB, Pouwer F. Are factor analytical techniques used appropriately in the validation of health status questionnaires? A systematic review on the quality of factor analysis of the SF-36. *Quality of life research*. 2005;14(5):1203-18.
109. de Vet HC, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM. Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. *Health and quality of life outcomes*. 2006;4(1):1-5.
110. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *Journal of clinical epidemiology*. 2003;56(5):395-407.
111. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *Journal of clinical epidemiology*. 2008;61(2):102-9.
112. Yost KJ, Eton DT, Garcia SF, Cella D. Minimally important differences were estimated for six Patient-Reported Outcomes Measurement Information System-Cancer scales in advanced-stage cancer patients. *Journal of clinical epidemiology*. 2011;64(5):507-16.

113. Embretson SE, Reise SP. Item response theory for psychologists. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers; 2000.
114. Terwee CB, Dekker FW, Wiersinga WM, Prummel MF, Bossuyt PMM. On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. *Quality of life research*. 2003;12(4):349-62.
115. Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. *Journal of chronic diseases*. 1986;39(11):897-906.
116. Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schünemann HJ. What is “quality of evidence” and why is it important to clinicians? *Bmj*. 2008;336(7651):995-8.
117. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *Bmj*. 2008;336(7650):924-6.
118. Siemieniuk R, Guyatt G. What is GRADE. *BMJ J Best Practice*. 2019;10.
119. Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. *Journal of clinical epidemiology*. 2011;64(4):401-6.
120. Langer G, Meerpohl JJ, Perleth M, Gartlehner G, Kaminski-Hartenthaler A, Schünemann H. GRADE guidelines: 1. Introduction–GRADE evidence

- profiles and summary of findings tables. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*. 2012;106(5):357-68.
121. Guyatt GH, Oxman AD, Kunz R, Atkins D, Brozek J, Vist G, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. *Journal of clinical epidemiology*. 2011;64(4):395-400.
122. Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, et al. GRADE guidelines 6. Rating the quality of evidence—imprecision. *Journal of clinical epidemiology*. 2011;64(12):1283-93.
123. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 7. Rating the quality of evidence—inconsistency. *Journal of clinical epidemiology*. 2011;64(12):1294-302.
124. Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, et al. GRADE guidelines: 5. Rating the quality of evidence—publication bias. *Journal of clinical epidemiology*. 2011;64(12):1277-82.
125. Guyatt GH, Oxman AD, Schünemann HJ, Tugwell P, Knottnerus A. GRADE guidelines: a new series of articles in the *Journal of Clinical Epidemiology*. *Journal of clinical epidemiology*. 2011;64(4):380-2.
126. Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, et al. GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias). *Journal of clinical epidemiology*. 2011;64(4):407-15.
127. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 8. Rating the quality of evidence—indirectness. *Journal of clinical epidemiology*. 2011;64(12):1303-10.

128. Guyatt GH, Oxman AD, Sultan S, Glasziou P, Akl EA, Alonso-Coello P, et al. GRADE guidelines: 9. Rating up the quality of evidence. *Journal of clinical epidemiology*. 2011;64(12):1311-6.
129. Mustafa RA, Santesso N, Brozek J, Akl EA, Walter SD, Norman G, et al. The GRADE approach is reproducible in assessing the quality of evidence of quantitative evidence syntheses. *Journal of clinical epidemiology*. 2013;66(7):736-42.
130. Odgaard-Jensen J, Vist GE, Timmer A, Kunz R, Akl EA, Schünemann H, et al. Randomisation to protect against selection bias in healthcare trials. *Cochrane database of systematic reviews*. 2011;(4).
131. Howick JH. *The philosophy of evidence-based medicine*. John Wiley & Sons; 2011.
132. Ioannidis JP. Why most published research findings are false. *PLoS medicine*. 2005;2(8):e124.
133. Guyatt GH, Oxman AD, Kunz R, Falck-Ytter Y, Vist GE, Liberati A, et al. Going from evidence to recommendations. *Bmj*. 2008;336(7652):1049-51.
134. Alonso-Coello P, Schünemann HJ, Moberg J, Brignardello-Petersen R, Akl EA, Davoli M, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction. *Gaceta sanitaria*. 2017;32(2):166-e1.
135. Alonso-Coello P, Oxman AD, Moberg J, Brignardello-Petersen R, Akl EA, Davoli M, et al. GRADE Evidence to Decision (EtD) frameworks: a

-
- systematic and transparent approach to making well informed healthcare choices. 2: Clinical practice guidelines. *Gaceta sanitaria*. 2017;32(2):167-e1.
136. de Dios FJG, Álvarez JCB, Rodríguez MG. Listas guía de comprobación de ensayos clínicos: declaración CONSORT. *Evidencias en pediatría*. 2011;7(3):22.
137. Moher D, Schulz KF, Altman DG, Group C. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. Elsevier; 2001.
138. González Rodríguez MP, Mayol V. Lista de comprobación de estudios sobre precisión de pruebas diagnósticas: declaración STARD. *Evid Pediatr*. 2012;8:43-43.
139. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Toward complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *American journal of clinical pathology*. 2003;119(1):18-22.
140. Manterola C, Otzen T. Estudios observacionales: los diseños utilizados con mayor frecuencia en investigación clínica. *International Journal of Morphology*. 2014;32(2):634-45.
141. Richason TP, Paulson SM, Lowenstein SR, Heard KJ. Case reports describing treatments in the emergency medicine literature: missing and misleading information. *BMC emergency medicine*. 2009;9(1):1-7.

142. Selman TJ, Morris RK, Zamora J, Khan KS. The quality of reporting of primary test accuracy studies in obstetrics and gynaecology: application of the STARD criteria. *BMC women's health*. 2011;11(1):1-7.
143. Thistlethwaite JE, Davies D, Ekeocha S, Kidd JM, MacDougall C, Matthews P, et al. The effectiveness of case-based learning in health professional education. A BEME systematic review: BEME Guide No. 23. *Medical teacher*. 2012;34(6):e421-44.
144. Vandembroucke JP, Von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *PLoS medicine*. 2007;4(10):e297.
145. Moraga J, Cartes-Velásquez R. Pautas de chequeo, parte I: CONSORT y TREND. *Revista chilena de cirugía*. 2015;67(2):225-32.
146. Vandembroucke JP, Von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, et al. Mejorar la comunicación de estudios observacionales en epidemiología (STROBE): explicación y elaboración. *Gaceta sanitaria*. 2009;23:158e1-28.
147. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandembroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Bulletin of the World Health Organization*. 2007;85:867-72.
148. Schriger DL. Suggestions for improving the reporting of clinical research: the role of narrative. *Annals of emergency medicine*. 2005;45(4):437-43.

149. The EQUATOR Network | Enhancing the QUALity and Transparency Of Health Research [Internet]. [citado 10 de diciembre de 2021]. Disponible en: <https://www.equator-network.org/>
150. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. Declaración de la Iniciativa STROBE (Strengthening the Reporting of Observational studies in Epidemiology): directrices para la comunicación de estudios observacionales. *Revista Española de Salud Pública*. 2008;82:251-9.
151. Cuschieri S. The STROBE guidelines. *Saudi journal of anaesthesia*. 2019;13(Suppl 1):S31.
152. STROBE [Internet]. STROBE. [citado 11 de diciembre de 2021]. Disponible en: <https://www.strobe-statement.org/>
153. Cuschieri S. WASP (Write a scientific paper): Understanding research metrics. *Early human development*. 2018;118:67-71.
154. Organization WH. Classification of diabetes mellitus. 2019;
155. Leslie RD, Palmer J, Schloot NC, Lernmark A. Diabetes at the crossroads: relevance of disease classification to pathophysiology and treatment. *Diabetologia*. 2016;59(1):13-20.
156. Organization WH. Report of a World Health Organization Consultation. Use of glycated haemoglobin (HbA1c) in the diagnosis of diabetes mellitus. *Diabetes Res Clin Pract*. 2011;93:299-309.
157. Organization WH. Global report on diabetes. Geneva: World Health Organization; 2016. ISBN 978-92-4-156525-7; 2016.

158. Roglic G, World Health Organization, editores. Global report on diabetes. Geneva, Switzerland: World Health Organization; 2016. 86 p.
159. Federation ID. IDF diabetes atlas 8th edition. International Diabetes Federation. 2017;905-11.
160. Tuomi T, Santoro N, Caprio S, Cai M, Weng J, Groop L. The many faces of diabetes: a disease with increasing heterogeneity. *The Lancet*. 2014;383(9922):1084-94.
161. Schwartz SS, Epstein S, Corkey BE, Grant SF, Gavin JR, Aguilar RB. The time is right for a new classification system for diabetes: rationale and implications of the β -cell-centric classification schema. *Diabetes care*. 2016;39(2):179-86.
162. Kahn SE, Cooper ME, Del Prato S. Pathophysiology and treatment of type 2 diabetes: perspectives on the past, present, and future. *The Lancet*. 2014;383(9922):1068-83.
163. Skyler JS, Bakris GL, Bonifacio E, Darsow T, Eckel RH, Groop L, et al. Differentiation of diabetes by pathophysiology, natural history, and prognosis. *Diabetes*. 2017;66(2):241-55.
164. Perl S, Kushner JA, Buchholz BA, Meeker AK, Stein GM, Hsieh M, et al. Significant human β -cell turnover is limited to the first three decades of life as determined by in vivo thymidine analog incorporation and radiocarbon dating. *The Journal of Clinical Endocrinology & Metabolism*. 2010;95(10):E234-9.

165. Hocking SL. Systematic review of the association between inherited genetic variants and response to blood glucose lowering therapies.
166. Zimmet P, Alberti K, Shaw J. Global and societal implications of the diabetes epidemic. *Nature*. 2001;414(6865):782-7.
167. Tan E, Khoo J, Gani LU, Malakar RD, Tay TL, Tirukonda PS, et al. Effect of multidisciplinary intensive targeted care in improving diabetes mellitus outcomes: a randomized controlled pilot study—the integrated diabetes education, awareness and lifestyle modification in Singapore (IDEALS) program. *Trials*. 2019;20(1):1-10.
168. Drincic AT, Akkireddy P, Knezevich JT. Common models used for inpatient diabetes management. *Current diabetes reports*. 2018;18(3):1-13.
169. Bansal V, Mottalib A, Pawar TK, Abbasakoor N, Chuang E, Chaudhry A, et al. Inpatient diabetes management by specialized diabetes team versus primary service team in non-critical care units: impact on 30-day readmission rate and hospital cost. *BMJ Open Diabetes Research and Care*. 2018;6(1):e000460.
170. Mandel SR, Langan S, Mathioudakis NN, Sidhaye AR, Bashura H, Bie JY, et al. Retrospective study of inpatient diabetes management service, length of stay and 30-day readmission rate of patients with diabetes at a community hospital. *Journal of community hospital internal medicine perspectives*. 2019;9(2):64-73.

171. Świętoniowska N, Sarzyńska K, Szymańska-Chabowska A, Jankowska-Polańska B. The role of education in type 2 diabetes treatment. *Diabetes research and clinical practice*. 2019;151:237-46.
172. Dyson PA, Twenefour D, Breen C, Duncan A, Elvin E, Goff L, et al. Diabetes UK evidence-based nutrition guidelines for the prevention and management of diabetes. *Diabetic medicine*. 2018;35(5):541-7.
173. Olatona FA, Airede CA, Aderibigbe SA, Osibogun A. Nutritional knowledge, dietary habits and nutritional status of diabetic patients attending teaching hospitals in Lagos, Nigeria. *Journal of Community Medicine and Primary Health Care*. 2019;31(2):90-103.
174. Maayeshi N, Mousavi SM, Ranjbaran H, Mirshekari M, Faghieh S. The relationship between nutritional knowledge and food habits and some cardiometabolic risk factors in patients with diabetes in Shiraz, Iran. *International Journal of Nutrition Sciences*. 2019;4(1):36-42.
175. Thewjitcharoen Y, Chotwanvirat P, Jantawan A, Siwasaranond N, Saetung S, Nimitphong H, et al. Evaluation of dietary intakes and nutritional knowledge in Thai patients with type 2 diabetes mellitus. *Journal of diabetes research*. 2018;2018.
176. Migliaccio S, Brasacchio C, Pivari F, Salzano C, Barrea L, Muscogiuri G, et al. What is the best diet for cardiovascular wellness? A comparison of different nutritional models. *International Journal of Obesity Supplements*. 2020;10(1):50-61.

177. Wu JH, Micha R, Mozaffarian D. Dietary fats and cardiometabolic disease: mechanisms and effects on risk factors and outcomes. *Nature Reviews Cardiology*. 2019;16(10):581-601.
178. Ohishi M. Hypertension with diabetes mellitus: physiology and pathology. *Hypertension research*. 2018;41(6):389-93.
179. Sinclair AJ, Dunning T, Dhatariya K, Experts an IG of, Sheu WH-H, Lin S-Y, et al. Clinical guidelines for type 1 diabetes mellitus with an emphasis on older adults: an executive summary. *Diabetic medicine*. 2020;37(1):53-70.
180. Aberle J, Lautenbach A, Meyhoefer S, Schmid SM, Selig L, Blüher M. Obesity and diabetes. *Experimental and Clinical Endocrinology & Diabetes*. 2021;129(S 01):S44-51.
181. Teich T, Zaharieva DP, Riddell MC. Advances in exercise, physical activity, and diabetes mellitus. *Diabetes technology & therapeutics*. 2019;21(S1):S-112.
182. Sigal RJ, Armstrong MJ, Bacon SL, Boule NG, Dasgupta K, Kenny GP, et al. Physical activity and diabetes. *Canadian journal of diabetes*. 2018;42:S54-63.
183. Blaslov K, Naran\ dja FS, Kruljac I, Renar IP. Treatment approach to type 2 diabetes: Past, present and future. *World journal of diabetes*. 2018;9(12):209.
184. Silver B, Ramaiya K, Andrew SB, Fredrick O, Bajaj S, Kalra S, et al. EADSG guidelines: insulin therapy in diabetes. *Diabetes Therapy*. 2018;9(2):449-92.
185. Mathew TK, Tadi P. Blood glucose monitoring. *StatPearls [Internet]*. 2020;
186. Mikhael EM, Hassali MA, Hussain SA. Effectiveness of diabetes self-management educational programs for type 2 diabetes mellitus patients in

- middle east countries: A systematic review. *Diabetes, metabolic syndrome and obesity: targets and therapy*. 2020;13:117.
187. Siminerio L, Hamm M, Kanter J, Cameron F de A, Krall J. A diabetes education model in primary care: provider and staff perspectives. *The Diabetes Educator*. 2019;45(5):498-506.
188. Nor NM, Shukri NAM, Yassin N, Sidek S, Azahari N. Barriers and enablers to make lifestyle changes among type 2 diabetes patients: a review. *Sains Malaysiana*. 2019;48(7):1491-502.
189. Uusitupa M, Khan TA, Viguioliouk E, Kahleova H, Rivellese AA, Hermansen K, et al. Prevention of type 2 diabetes by lifestyle changes: A systematic review and meta-analysis. *Nutrients*. 2019;11(11):2611.
190. Schmidt SK, Hemmestad L, MacDonald CS, Langberg H, Valentiner LS. Motivation and barriers to maintaining lifestyle changes in patients with type 2 diabetes after an intensive lifestyle intervention (The U-TURN Trial): a longitudinal qualitative study. *International Journal of Environmental Research and Public Health*. 2020;17(20):7454.
191. Cáncer [Internet]. [citado 5 de enero de 2022]. Disponible en: <https://www.who.int/es/news-room/fact-sheets/detail/cancer>
192. Ferlay J, Colombet M, Soerjomataram I, Parkin DM, Piñeros M, Znaor A, et al. Cancer statistics for the year 2020: An overview. *International Journal of Cancer*. 2021;

193. Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, et al. Global cancer observatory: cancer today. Lyon: International Agency for Research on Cancer; 2018. 2020.
194. Islami F, Goding Sauer A, Miller KD, Siegel RL, Fedewa SA, Jacobs EJ, et al. Proportion and number of cancer cases and deaths attributable to potentially modifiable risk factors in the United States. *CA: a cancer journal for clinicians*. 2018;68(1):31-54.
195. Mattiuzzi C, Lippi G. Current cancer epidemiology. *Journal of epidemiology and global health*. 2019;9(4):217.
196. Wild CP, Espina C, Bauld L, Bonanni B, Brenner H, Brown K, et al. Cancer prevention Europe. *Molecular oncology*. 2019;13(3):528-34.
197. de Martel C, Georges D, Bray F, Ferlay J, Clifford GM. Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis. *The Lancet Global Health*. 2020;8(2):e180-90.
198. Loud JT, Murphy J. Cancer screening and early detection in the 21st century. En: *Seminars in oncology nursing*. Elsevier; 2017. p. 121-8.
199. Gaikwad PS, Banerjee R. Advances in point-of-care diagnostic devices in cancers. *Analyst*. 2018;143(6):1326-48.
200. Hawkes N. Cancer survival data emphasise importance of early diagnosis. *British Medical Journal Publishing Group*; 2019.
201. Smith RA, Andrews KS, Brooks D, Fedewa SA, Manassaram-Baptiste D, Saslow D, et al. Cancer screening in the United States, 2018: a review of

- current American Cancer Society guidelines and current issues in cancer screening. *CA: a cancer journal for clinicians*. 2018;68(4):297-316.
202. William W, Ware A, Basaza-Ejiri AH, Obungoloch J. A review of image analysis and machine learning techniques for automated cervical cancer screening from pap-smear images. *Computer methods and programs in biomedicine*. 2018;164:15-22.
203. Seely JM, Alhassan T. Screening for breast cancer in 2018—what should we be doing today? *Current Oncology*. 2018;25(s1):115-24.
204. Schünemann HJ, Lerda D, Quinn C, Follmann M, Alonso-Coello P, Rossi PG, et al. Breast cancer screening and diagnosis: a synopsis of the European Breast Guidelines. *Annals of internal medicine*. 2020;172(1):46-56.
205. Bevan R, Rutter MD. Colorectal cancer screening—who, how, and when? *Clinical endoscopy*. 2018;51(1):37.
206. Dickens E, Ahmed S. Principles of cancer treatment by chemotherapy. *Surgery (Oxford)*. 2018;36(3):134-8.
207. Wang Y, Deng W, Li N, Neri S, Sharma A, Jiang W, et al. Combining immunotherapy and radiotherapy for cancer treatment: current challenges and future directions. *Frontiers in pharmacology*. 2018;9:185.
208. O'Donnell JS, Hoefsmit EP, Smyth MJ, Blank CU, Teng MW. The promise of neoadjuvant immunotherapy and surgery for cancer treatment. *Clinical Cancer Research*. 2019;25(19):5743-51.

-
209. Billings JA. What is palliative care? *Journal of palliative medicine*. 1998;1(1):73-81.
210. Kvale PA, Simoff M, Prakash UB. Palliative care. *Chest*. 2003;123(1):284S-311S.
211. Levy MH, Adolph MD, Back A, Block S, Codada SN, Dalal S, et al. Palliative care. *Journal of the National Comprehensive Cancer Network*. 2012;10(10):1284-309.
212. Morrison RS, Meier DE. Palliative care. *New England Journal of Medicine*. 2004;350(25):2582-90.
213. Sepúlveda C, Marlin A, Yoshida T, Ullrich A. Palliative care: the World Health Organization's global perspective. *Journal of pain and symptom management*. 2002;24(2):91-6.
214. Twycross RG. *Introducing palliative care*. Radcliffe Publishing; 2003.
215. Hui D, Bruera E. Models of palliative care delivery for patients with cancer. *Journal of Clinical Oncology*. 2020;38(9):852.
216. Martínez AC, Cortés CL, Rodríguez LV, Donoso AM, Sarto LE, Linés MPC. Cuidados proactivos en el paciente al final de la vida. cuidados paliativos en domicilio. *Revista Sanitaria de Investigación*. 2021;2(4):35.
217. Knaul FM, Bhadelia A, Rodríguez NM, Arreola-Ornelas H, Zimmermann C. The Lancet Commission on Palliative Care and Pain Relief—findings, recommendations, and future directions. *The Lancet Global Health*. 2018;6:S5-6.

218. Rachels JA. Active and passive euthanasia. Duke University Press; 2019.
219. Rodea AS, Sanjuan MF. Tratamientos al final de la vida: cuidados paliativos, sedación terminal, eutanasia y suicidio medicamente asistido (SMA). Revista de Investigación y Educación en Ciencias de la Salud (RIECS). 2021;6(2):94-105.
220. Fernández MO. La legalización de la eutanasia activa directa en España: la Ley Orgánica 3/2021, de 24 de marzo, de regulación de la eutanasia. Diario La Ley. 2021;(9830):4.
221. Jefatura del Estado. Ley Orgánica 3/2021, de 24 de marzo, de regulación de la eutanasia [Internet]. Sec. 1, Ley Orgánica 3/2021 mar 25, 2021 p. 34037-49. Disponible en: <https://www.boe.es/eli/es/lo/2021/03/24/3>
222. Ayres AJ. La integración sensorial en los niños: desafíos sensoriales ocultos. Tea; 2008.
223. del Moral Orro G, Montañó MÁP, Valer PS. Del marco teórico de integración sensorial al modelo clínico de intervención. Revista electrónica de terapia ocupacional Galicia, TOG. 2013;(17):18.
224. Serrano P. La integración sensorial: en el desarrollo y aprendizaje infantil. Vol. 85. Narcea ediciones; 2019.
225. Smith MC. Sensory integration: Theory and practice. FA Davis; 2019.
226. Miller LJ, Anzalone ME, Lane SJ, Cermak SA, Osten ET. Concept evolution in sensory integration: a proposed nosology for diagnosis. Am J Occup Ther. abril de 2007;61(2):135-40.

227. Schaaf R, Benevides T, Blanche E, Brett-Green B, Burke J, Cohn E, et al. Parasympathetic functions in children with sensory processing disorder. *Frontiers in Integrative Neuroscience* [Internet]. 2010 [citado 10 de enero de 2022];4:4. Disponible en: <https://www.frontiersin.org/article/10.3389/fnint.2010.00004>
228. Bar-Shalita T, Granovsky Y, Parush S, Weissman-Fogel I. Sensory modulation disorder (SMD) and pain: a new perspective. *Frontiers in integrative neuroscience*. 2019;13:27.
229. Adra N, Cao A, Makris N, Valera EM. Sensory Modulation Disorder and its Neural Circuitry in Adults with ADHD: A Pilot Study. *Brain Imaging and Behavior*. 2021;15(2):930-40.
230. Mazor-Karsenty T, Parush S, Shalev L. Sustained attention in sensory modulation disorder and attention deficit/hyperactivity disorder. *Research in developmental disabilities*. 2019;88:22-9.
231. Machingura T, Shum D, Molineux M, Lloyd C. Effectiveness of sensory modulation in treating sensory modulation disorders in adults with schizophrenia: a systematic literature review. *International Journal of Mental Health and Addiction*. 2018;16(3):764-80.
232. Bair W-N, Kiemel T, Jeka JJ, Clark JE. Development of Multisensory Reweighting Is Impaired for Quiet Stance Control in Children with Developmental Coordination Disorder (DCD). *PLOS ONE* [Internet]. 18 de julio de 2012 [citado 10 de enero de 2022];7(7):e40932. Disponible en: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0040932>

-
233. Hashemi A, Shahrbanian S. The effects of sensory-motor integration on gross motor function in children with developmental coordination disorder. *Journal of Sabzevar University of Medical Sciences*. 2018;25(3):419-27.
234. Rose ASRAS, Basit H. Effects of Sensorimotor Problems on the Performance of Activities of Daily Living in Children with Autism Spectrum Disorder. 2020;
235. Lonkar H. An Overview of Sensory Processing Disorder. :36.
236. Jean Ayres A. *La integración sensorial y el niño*. Sevilla: MAD. 2006;
237. Beaudry I. *Problemas de aprendizaje en la infancia. La descoordinación motriz, la hiperactividad y las dificultades académicas desde el enfoque de la teoría de la integración sensorial*. Oviedo: Ediciones Nobel; 2011.
238. Kilroy E, Aziz-Zadeh L, Cermak S. Ayres theories of autism and sensory integration revisited: What contemporary neuroscience has to say. *Brain sciences*. 2019;9(3):68.
239. Guardado KE, Sergent SR. *Sensory Integration*. StatPearls [Internet]. 2021;
240. May-Benson TA, de Mello Gentil JL, Teasdale A. Characteristics and prevalence of gravitational insecurity in children with sensory processing dysfunction. *Research in developmental disabilities*. 2020;101:103640.
241. Dumas M, Knox R, O'Brien C, Craig CE. Dissociating proprioceptive deficits in Autism Spectrum Disorders: Intact acuity but impaired sensory integration in postural control. *bioRxiv*. 2019;644617.

-
242. Lane SJ, Mailloux Z, Schoen S, Bundy A, May-Benson TA, Parham LD, et al. Neural Foundations of Ayres Sensory Integration®. *Brain sciences*. 2019;9(7):153.
243. Peterka RJ. Sensory integration for human balance control. *Handbook of clinical neurology*. 2018;159:27-42.
244. Mailloux Z, Parham LD, Roley SS, Ruzzano L, Schaaf RC. Introduction to the evaluation in ayres sensory integration®(EASI). *American Journal of Occupational Therapy*. 2018;72(1):7201195030p1-7.
245. Hatch-Rasmussen C. Sensory integration. Center for the Study of Autism at www.autism.org/si.html. 1995;
246. Smith Roley S, Mailloux Z, Miller-Kuhaneck H, Glennon T. Understanding Ayres' sensory integration. 2007;
247. Ayres AJ, Mailloux ZK, Wendler CL. Developmental dyspraxia: Is it a unitary function? *The occupational Therapy journal of Research*. 1987;7(2):93-110.
248. Bodison SC. Developmental dyspraxia and the play skills of children with autism. *American Journal of Occupational Therapy*. 2015;69(5):6905185060p1-6.
249. Cummins RA. Sensory integration and learning disabilities: Ayres' factor analyses reappraised. *Journal of Learning Disabilities*. 1991;24(3):160-8.
250. Brown C, Dunn W. Adult/adolescent sensory profile: User's manual. San Antonio, Texas: Psychological Corporation. 2002;

-
251. Dunn W. Sensory profile: User's manual. Psychological Corporation San Antonio, TX; 1999.
252. Dunn W. The impact of sensory processing abilities on the daily lives of young children and their families: A conceptual model. *Infants and young children*. 1997;9:23-35.
253. Ávila Álvarez A, Martínez Piédrola R, Matilla Mora R, Máximo Bocanegra M, Méndez Méndez B, Talavera Valverde MA. Marco de Trabajo para la práctica de la Terapia Ocupacional: Dominio y proceso. 2da Edición [Traducción]. AOTA Recuperado de www.terapia-ocupacional.com/aota2010esp.pdf. 2008;
254. Schoen SA, Lane SJ, Mailloux Z, May-Benson T, Parham LD, Smith Roley S, et al. A systematic review of ayres sensory integration intervention for children with autism. *Autism Research*. 2019;12(1):6-19.
255. Schaaf RC, Dumont RL, Arbesman M, May-Benson TA. Efficacy of occupational therapy using Ayres Sensory Integration®: A systematic review. *American Journal of Occupational Therapy*. 2018;72(1):7201190010p1-10.
256. Parham LD, Roley SS, May-Benson TA, Koomar J, Brett-Green B, Burke JP, et al. Development of a fidelity measure for research on the effectiveness of the Ayres Sensory Integration® intervention. *American Journal of Occupational Therapy*. 2011;65(2):133-42.
257. Watling R, Hauer S. Effectiveness of Ayres Sensory Integration® and sensory-based interventions for people with autism spectrum disorder: A

-
- systematic review. *American journal of occupational therapy*. 2015;69(5):6905180030p1-12.
258. Haywood KL, Garratt AM, Fitzpatrick R. Quality of life in older people: a structured review of self-assessed health instruments. *Expert Review of Pharmacoeconomics & Outcomes Research*. 2006;6(2):181-94.
259. Alla S, Sullivan SJ, Hale L, McCrory P. Self-report scales/checklists for the measurement of concussion symptoms: a systematic review. *British journal of sports medicine*. 2009;43(Suppl 1):i3-12.
260. Marinus J, Ramaker C, van Hilten JJ, Stiggelbout AM. Health related quality of life in Parkinson's disease: a systematic review of disease specific instruments. *Journal of Neurology, Neurosurgery & Psychiatry*. 2002;72(2):241-8.
261. Wind H, Gouttebarga V, Kuijer P, Frings-Dresen MH. Assessment of functional capacity of the musculoskeletal system in the context of work, daily living, and sport: a systematic review. *Journal of Occupational Rehabilitation*. 2005;15(2):253-72.
262. Ioannidis JP, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, et al. Increasing value and reducing waste in research design, conduct, and analysis. *The Lancet*. 2014;383(9912):166-75.
263. Agency EM. Committee for medicinal products for human use. Scientific Discussion: Taxotere (Docetaxel). European Medicines Agency London, UK; 2005.

-
264. Marshall M, Lockwood A, Bradley C, Adams C, Joy C, Fenton M. Unpublished rating scales: a major source of bias in randomised controlled trials of treatments for schizophrenia. *The British Journal of Psychiatry*. 2000;176(3):249-52.
265. Terwee CB, Prinsen CA, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Quality of Life Research*. 2018;27(5):1159-70.
266. Sijtsma K. On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*. 2009;74(1):107-20.
267. Cortina JM. What is coefficient alpha? An examination of theory and applications. *Journal of applied psychology*. 1993;78(1):98.
268. De Leeuw ED, Hox J, Dillman D. *International handbook of survey methodology*. Routledge; 2012.
269. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin*. 1979;86(2):420.
270. Cohen J. Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit. *Psychological bulletin*. 1968;70(4):213.
271. de Vet HC, Mokkink LB, Terwee CB, Hoekstra OS, Knol DL. Clinicians are right not to like Cohen's κ . *Bmj*. 2013;346.

-
272. Prinsen CA, Mokkink LB, Bouter LM, Alonso J, Patrick D, de Vet H, et al. COSMIN guideline for systematic reviews of outcome measurement instruments. *Quality of Life Research*. 2016;25:21-21.
273. Messick S. Test validity and the ethics of assessment. *American psychologist*. 1980;35(11):1012.
274. Patrick DL, Burke LB, Gwaltney CJ, Leidy NK, Martin ML, Molsen E, et al. Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: part 2—assessing respondent understanding. *Value in Health*. 2011;14(8):978-88.
275. Patrick DL, Burke LB, Gwaltney CJ, Leidy NK, Martin ML, Molsen E, et al. Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: part 1—eliciting concepts for a new PRO instrument. *Value in Health*. 2011;14(8):967-77.
276. Brod M, Tesler LE, Christensen TL. Qualitative research and content validity: developing best practices based on science and experience. *Quality of life research*. 2009;18(9):1263-78.
277. Association AP, Education NC on M in. American educational research association. Washington, DC: American Psychological Association: Standards for Educational and Psychological Testing. 2014;

-
278. Chiarotto A, Ostelo RW, Boers M, Terwee CB. A systematic review highlights the need to investigate the content validity of patient-reported outcome measures for physical functioning in patients with low back pain. *Journal of clinical epidemiology*. 2018;95:73-93.
279. Lohr KN, Aaronson NK, Alonso J, Burnam MA, Patrick DL, Perrin EB, et al. Evaluating quality-of-life and health status instruments: development of scientific review criteria. *Clinical therapeutics*. 1996;18(5):979-92.
280. Reeve BB, Wyrwich KW, Wu AW, Velikova G, Terwee CB, Snyder CF, et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Quality of Life Research*. 2013;22(8):1889-905.
281. Chou R, Baker WL, Bañez LL, Iyer S, Myers ER, Newberry S, et al. Agency for Healthcare Research and Quality Evidence-based Practice Center methods provide guidance on prioritization and selection of harms in systematic reviews. *Journal of Clinical Epidemiology*. 2018;98:98-104.
282. Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD, et al. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. *Jama*. 2013;309(8):814-22.
283. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *International journal of nursing studies*. 2011;48(6):661-71.

284. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of internal medicine*. 2011;155(8):529-36.
285. Lucas NP, Macaskill P, Irwig L, Bogduk N. The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). *Journal of clinical epidemiology*. 2010;63(8):854-61.
286. McDowell I, Jenkinson C. Development standards for health measures. *Journal of health services research & policy*. 1996;1(4):238-46.
287. Andresen EM. Criteria for assessing the tools of disability outcomes research. *Archives of physical medicine and rehabilitation*. 2000;81:S15-20.
288. Bombardier C, Tugwell P. Methodological considerations in functional assessment. *The journal of rheumatology Supplement*. 1987;14:6-10.
289. Bot SDM, Terwee CB, Van der Windt D, Bouter LM, Dekker J, De Vet HCW. Clinimetric evaluation of shoulder disability questionnaires: a systematic review of the literature. *Annals of the rheumatic diseases*. 2004;63(4):335-41.
290. De Boer MR, Moll AC, De Vet HC, Terwee CB, Völker-Dieben HJ, Van Rens GH. Psychometric properties of vision-related quality of life questionnaires: a systematic review. *Ophthalmic and Physiological Optics*. 2004;24(4):257-73.
291. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJM, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Controlled clinical trials*. 1996;17(1):1-12.

-
292. L'insalata JC, Warren RF, Cohen SB, Altchek DW, Peterson MG. A self-administered questionnaire for assessment of symptoms and function of the shoulder. *JBJS*. 1997;79(5):738-48.
293. Strauss ME. Introduction to the special section on construct validity of psychological tests: 50 years after Cronbach and Meehl (1955). *Psychological Assessment*. 2005;17(4):395.
294. Stratford PW, Riddle DL. Assessing sensitivity to change: choosing the appropriate change coefficient. *Health and Quality of Life Outcomes*. 2005;3(1):1-7.
295. Zou GY. Quantifying responsiveness of quality of life measures without an external criterion. *Quality of life research*. 2005;14(6):1545-52.
296. Beaton DE, Hogg-Johnson S, Bombardier C. Evaluating changes in health status: reliability and responsiveness of five generic health status measures in workers with musculoskeletal disorders. *Journal of clinical epidemiology*. 1997;50(1):79-93.
297. Stucki G, Liang MH, Fossel AH, Katz JN. Relative responsiveness of condition-specific and generic health status measures in degenerative lumbar spinal stenosis. *Journal of clinical epidemiology*. 1995;48(11):1369-78.
298. Wright JG, Young NL. A comparison of different indices of responsiveness. *Journal of clinical epidemiology*. 1997;50(3):239-46.

299. Sackett DL, Rosenberg WM, Gray JM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. Vol. 312, Bmj. British Medical Journal Publishing Group; 1996. p. 71-2.
300. Boulet L-P, Becker A, Bérubé D, Beveridge R, Ernst P. Canadian asthma consensus report, 1999. Cmaj. 1999;161(11 suppl 1):S1-5.
301. Ernst P, Fitzgerald JM, Spier S. Canadian asthma consensus conference summary of recommendations. Canadian Respiratory Journal. 1996;3(2):89-101.
302. J Jovella A, Navarro-Rubiob MD. Evaluación de la evidencia científica. 1995;
303. Niveles de Evidencia [Internet]. [citado 8 de febrero de 2022]. Disponible en: https://www.sspa.juntadeandalucia.es/servicioandaluzdesalud/hrs3/index.php?id=manual_procedimientos_niveles
304. Ebell MH, Siwek J, Weiss BD, Woolf SH, Susman J, Ewigman B, et al. Strength of recommendation taxonomy (SORT): a patient-centered approach to grading evidence in the medical literature. The Journal of the American Board of Family Practice. 2004;17(1):59-67.
305. Niveles de evidencia científica en NANDA-I 2021-2023 [Internet]. [citado 8 de febrero de 2022]. Disponible en: <http://www.eldiagnosticoenfermero.es/2021/03/niveles-de-evidencia-cientifica-en.html>
306. Merlin T, Weston A, Tooher R. Extending an evidence hierarchy to include topics other than treatment: revising the Australian 'levels of evidence'. BMC medical research methodology. 2009;9(1):1-8.

-
307. Howick J, Phillips B, Ball C, Sackett D, Badenoch D, Straus S. Oxford Centre for Evidence-Based Medicine (CEBM). Levels of evidence and grades of recommendation (March 2009).[Internet]. Oxford: CEBM; 2009.
308. NICE | The National Institute for Health and Care Excellence [Internet]. NICE. NICE; [citado 8 de febrero de 2022]. Disponible en: <https://www.nice.org.uk/>
309. PCCRP - Practicing Chiropractors' Committee on Radiology Protocols [Internet]. [citado 8 de febrero de 2022]. Disponible en: <http://www.pccrp.org/>
310. Sign S. 50: A guideline developers' handbook. SIGN. 2011;
311. Birtwhistle R, Pottie K, Shaw E, Dickinson JA, Brauer P, Fortin M, et al. Canadian task force on preventive health care: We're back! Canadian Family Physician. 2012;58(1):13-5.
312. United States Preventive Services Taskforce [Internet]. [citado 8 de febrero de 2022]. Disponible en: <https://www.uspreventiveservicestaskforce.org/uspstf/grades.htm>
313. Association AD. Introduction. ADA evidence-grading system for clinical practice recommendations. Diabetes Care. 2008;31:S1-2.
314. Greenland P, Alpert JS. Practice Guidelines 2010 ACCF/AHA Guideline for Assessment of Cardiovascular Risk in Asymptomatic Adults. Circulation. 2010;122:e584-636.

-
315. Granholm A, Alhazzani W, Møller MH. Use of the GRADE approach in systematic reviews and guidelines. *British journal of anaesthesia*. 2019;123(5):554-9.
316. Gándara-Gafo B, Santos-del Riego S, Viana-Moldes I, Muñiz J. Cultural Adaptation of the Adolescent/Adult Sensory Profile for Spain. *American Journal of Occupational Therapy*. 2019;73(6):7306205070p1-9.
317. Gándara-Gafo B, Santos-del Riego S, Muñiz J. Reference Values for the Adolescent/Adult Sensory Profile in Spain. *American Journal of Occupational Therapy*. 2019;73(5):7305205040p1-8.
318. Magasi S, Ryan G, Revicki D, Lenderking W, Hays RD, Brod M, et al. Content validity of patient-reported outcome measures: perspectives from a PROMIS meeting. *Quality of Life Research*. 2012;21(5):739-46.
319. Leidy NK, Vernon M. Perspectives on patient-reported outcomes. *Pharmacoeconomics*. 2008;26(5):363-70.
320. Rothman M, Burke L, Erickson P, Leidy NK, Patrick DL, Petrie CD. Use of existing patient-reported outcome (PRO) instruments and their modification: the ISPOR Good Research Practices for Evaluating and Documenting Content Validity for the Use of Existing Instruments and Their Modification PRO Task Force Report. *Value in Health*. 2009;12(8):1075-83.
321. Floyd FJ, Widaman KF. Factor analysis in the development and refinement of clinical assessment instruments. *Psychological assessment*. 1995;7(3):286.

-
322. Gothwal S, Mohan A, Khan F, Om P. Comparison of Major Complication Rate in High and Low Surgical Apgar Score in Abdominal Surgery Cases. *Indian Journal of Surgery*. 2019;81(2):164-8.
323. Fayers PM, Hand DJ. Causal variables, indicator variables and measurement scales: an example from quality of life. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. 2002;165(2):233-53.
324. Streiner DL. Being inconsistent about consistency: When coefficient alpha does and doesn't matter. *Journal of personality assessment*. 2003;80(3):217-22.
325. Streiner DL. Figuring out factors: the use and misuse of factor analysis. *The Canadian Journal of Psychiatry*. 1994;39(3):135-40.
326. Bollen KA. *Structural equations with latent variables*. Vol. 210. John Wiley & Sons; 1989.
327. Nunnally JC. *Psychometric theory* 3E. Tata McGraw-hill education; 1994.
328. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychological methods*. 1996;1(1):30.
329. Mokkink LB. COSMIN risk of bias checklist. *Amsterdam Public Heal. Res. Inst.*; 2018.
330. Messick S. The standard problem: Meaning and values in measurement and evaluation. *American psychologist*. 1975;30(10):955.

-
331. Altman DG. Practical statistics for medical research Chapman and Hall. London and New York. 1991;
332. De Boer MR, Terwee CB, de Vet HC, Moll AC, Völker-Dieben HJ, van Rens GH. Evaluation of cross-sectional and longitudinal construct validity of two vision-related quality of life questionnaires: the LVQOL and VCM1. *Quality of Life Research*. 2006;15(2):233-48.
333. Cohen J. Statistical power analysis for the behavioural sciences 2 edition Lawrence Erlbaum Associates. Hillsdale, NJ. 1988;
334. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Quality of life research*. 1995;4(4):293-307.
335. Norman GR. Issues in the use of change scores in randomized trials. *Journal of clinical epidemiology*. 1989;42(11):1097-105.
336. Stockler MR, Osoba D, Goodwin P, Corey P, Tannock IF. Responsiveness to change in health-related quality of life in a randomized clinical trial: a comparison of the Prostate Cancer Specific Quality of Life Instrument (PROSQOLI) with analogous scales from the EORTC QLQ-C30 and a trial specific module. *Journal of Clinical Epidemiology*. 1998;51(2):137-45.
337. Chen X-L, Zhong L, Wen Y, Liu T-W, Li X-Y, Hou Z-K, et al. Inflammatory bowel disease-specific health-related quality of life instruments: a systematic review of measurement properties. *Health and quality of life outcomes*. 2017;15(1):1-13.

-
338. Sierevelt IN, Zwiers R, Schats W, Haverkamp D, Terwee CB, Nolte PA, et al. Measurement properties of the most commonly used Foot-and Ankle-Specific Questionnaires: the FFI, FAOS and FAAM. A systematic review. *Knee Surgery, Sports Traumatology, Arthroscopy*. 2018;26(7):2059-73.
339. Gagnier JJ, Mullins M, Huang H, Marinac-Dabic D, Ghambaryan A, Eloff B, et al. A systematic review of measurement properties of patient-reported outcome measures used in patients undergoing total knee arthroplasty. *The Journal of Arthroplasty*. 2017;32(5):1688-97.
340. Treanor C, Donnelly M. A methodological review of the Short Form Health Survey 36 (SF-36) and its derivatives among breast cancer survivors. *Quality of life research*. 2015;24(2):339-62.
341. Ertzgaard P, Nene A, Kiekens C, Burns AS. A review and evaluation of patient-reported outcome measures for spasticity in persons with spinal cord damage: Recommendations from the Ability Network—an international initiative. *The Journal of Spinal Cord Medicine*. 2020;43(6):813-23.
342. Marti C, Hensler S, Herren DB, Niedermann K, Marks M. Measurement properties of the EuroQoL EQ-5D-5L to assess quality of life in patients undergoing carpal tunnel release. *Journal of Hand Surgery (European Volume)*. 2016;41(9):957-62.
343. Qian X, Tan RL-Y, Chuang L-H, Luo N. Measurement properties of commonly used generic preference-based measures in East and South-East Asia: a systematic review. *Pharmacoeconomics*. 2020;38(2):159-70.

344. Eyles JP, Hunter DJ, Meneses SRF, Collins NJ, Dobson F, Lucas BR, et al. Instruments assessing attitudes toward or capability regarding self-management of osteoarthritis: a systematic review of measurement properties. *Osteoarthritis and Cartilage*. 2017;25(8):1210-22.
345. Janssen CA, Oude Voshaar MA, Ten Klooster PM, Jansen TLTA, Vonkeman HE, van de Laar MA. A systematic literature review of patient-reported outcome measures used in gout: an evaluation of their content and measurement properties. *Health and quality of life outcomes*. 2019;17(1):1-11.
346. Whynes DK, McCahon RA, Ravenscroft A, Hodgkinson V, Evley R, Hardman JG. Responsiveness of the EQ-5D health-related quality-of-life instrument in assessing low back pain. *Value in Health*. 2013;16(1):124-32.
347. Mason SJ, Catto JW, Downing A, Bottomley SE, Glaser AW, Wright P. Evaluating patient-reported outcome measures (PROMs) for bladder cancer: a systematic review using the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) checklist. *BJU international*. 2018;122(5):760.
348. Szekeres M. Clinical relevance commentary in response to: The validity and clinical utility of the Disabilities of the Arm Shoulder and Hand questionnaire for hand injuries in developing country contexts: A systematic review. *Journal of hand therapy: official journal of the American Society of Hand Therapists*. 2018;31(1):91-2.
349. Mokkink LB, Terwee CB, Gibbons E, Stratford PW, Alonso J, Patrick DL, et al. Inter-rater agreement and reliability of the COSMIN (COnsensus-based

- Standards for the selection of health status Measurement Instruments) checklist. *BMC Medical Research Methodology*. 2010;10(1):1-11.
350. Schünemann HJ, Fretheim A, Oxman AD. Improving the use of research evidence in guideline development: 1. Guidelines for guidelines. *Health Research Policy and Systems*. 2006;4(1):1-6.
351. Schünemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Vist GE, et al. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *Bmj*. 2008;336(7653):1106-10.
352. Schünemann HJ, Fretheim A, Oxman AD. Improving the use of research evidence in guideline development: 9. Grading evidence and recommendations. *Health Research Policy and Systems*. 2006;4(1):1-7.

ANEXOS

ANEXO I. COMITÉ DE ÉTICA UCAM



COMITÉ DE ÉTICA DE LA UCAM

DATOS DEL PROYECTO

Título:	"Síntomatología sensorial en pacientes españoles supervivientes de COVID-19"	
Investigador Principal	Nombre	Correo-e
Dr.	David Pérez Cruzado	dpacruzado@ucam.edu

INFORME DEL COMITÉ

Fecha	30/10/2020	Código	CE102003
--------------	------------	---------------	----------

Tipo de Experimentación

Investigación experimental clínica con seres humanos	
Investigación experimental no clínica con seres humanos	
Utilización de tejidos humanos procedentes de pacientes, personas sanas, tejidos embrionarios o fetales	
Utilización de tejidos humanos, tejidos embrionarios o fetales procedentes de bancos de muestras o tejidos	
Investigación observacional con seres humanos, psicológica o comportamental en humanos	
Uso de datos personales, información genética, etc.	X
Experimentación animal	
Utilización de agentes biológicos de riesgo para la salud humana, animal o las plantas	
Uso de organismos modificados genéticamente (OMGs)	

Comentarios Respecto al Tipo de Experimentación
Nada Obsta

Comentarios Respecto a la Metodología de Experimentación
Nada Obsta





COMITÉ DE ÉTICA DE LA UCAM

Sugerencias al Investigador

--

A la vista de la solicitud de informe adjunto por el Investigador y de las recomendaciones anteriormente expuestas el dictamen del Comité es:

Emitir Informe Favorable	X
Emitir Informe Desfavorable	
Emitir Informe Favorable condicionado a Subsanación	

MOTIVACIÓN

Incrementará conocimientos en su área

Vº Bº El Presidente,

Fdo.: José Alberto Cánovas Sánchez



El Secretario,

Fdo.: José Alarcón Teruel



A/A Comité de Ética de la Investigación UCAM

Yo, David Pérez Cruzado, con DNI 76637145H e IP del proyecto titulado "Síntomatología sensorial en pacientes españoles supervivientes de COVID-19" con código CE102003 (proyecto de investigación relacionado con el COVID-19. Plan propio de apoyo a la I+D+I 2019-2020 de la Universidad Católica de Murcia. PMAFI-COVID19/08). Comunico la inclusión del investigador Pedro Montagut Martínez con DNI 23267766T con el fin de aumentar el equipo investigador en dicho proyecto, enfocándose en la elaboración de documentos científicos, interpretación de resultados así como difusión de estos. Además de ello, dicho proyecto, formará parte de la tesis Doctoral que el investigador Pedro Montagut Martínez está realizando en dicha Universidad bajo mi tutorización.

Fd: David Pérez Cruzado

A handwritten signature in black ink, appearing to be "D. Pérez Cruzado", is written below the typed name.

A 23 de Septiembre de 2021

ANEXO II. CUESTIONARIO AASP- 42 ÍTEMS



PERFIL SENSORIAL

ADOLESCENTES Y ADULTOS

(42 Ítems)

Auto Cuestionario

Nombre: _____ Edad: _____ Fecha: _____

Fecha de nacimiento: _____ Género: masculino femenino

¿Hay aspectos de su vida diaria que no lo satisfacen? Si es sí, por favor explíquelos _____

INSTRUCCIONES

Por favor marque el cuadro que **mejor** describa la frecuencia con la cual usted realiza los siguientes comportamientos. Si usted no lo sabe ya que no tiene experiencia en ese comportamiento en particular, por favor dibuje una X sobre el número del ítem. Puede escribir cualquier comentario al final de cada sección.

Por favor responda todos los enunciados. Utilice las siguientes referencias para marcar su respuesta

CASI NUNCA	Cuando se presenta la oportunidad, usted casi nunca responde de esa manera (aproximadamente el 5% o menos del tiempo).
RARAMENTE	Cuando se presenta la oportunidad, usted raramente responde de esa manera (aproximadamente el 25% del tiempo).
A VECES	Cuando se presenta la oportunidad, usted a veces responde de esa manera (aproximadamente el 50% del tiempo).
FRECUENTEMENTE	Cuando se presenta la oportunidad, usted frecuentemente responde de esa manera (aproximadamente el 75% del tiempo).
CASI SIEMPRE	Cuando se presenta la oportunidad, usted casi siempre responde de esa manera (aproximadamente el 95% o más del tiempo).

ITEM		A. Procesamiento Gusto/Olfato	Casi nunca	Raramente	A veces	Frecuentemente	Casi siempre
	1	Añado condimentos/especias a mi comida.					
	2	Disfruto cuando estoy cerca de las personas que usan perfumes o lociones.					
	3	Muchos alimentos no tienen sabor para mí (en otras palabras, no tienen mucho gusto o son insulsos).					
	4	Voy a oler flores frescas cuando las veo.					

Comentarios:

ITEM		B. Procesamiento Movimiento	Casi nunca	Raramente	A veces	Frecuentemente	Casi siempre
	5	Tengo miedo a las alturas.					
	6	Disfruto como me siento cuando me muevo (por ejemplo; bailando, corriendo).					
	7	Evito los ascensores y/o las escaleras mecánicas porque no me gusta el movimiento.					
	8	Tropiezo o me choco con las cosas.					
	9	Elijo participar en actividades físicas.					
	10	No me siento seguro cuando camino por escaleras (por ejemplo; me tropiezo, pierdo el equilibrio y/o necesito sujetarme al pasamano).					
	11	Me mareo fácilmente (por ejemplo; después de inclinarme, levantarme rápido).					

Comentarios:

ITEM		C. Procesamiento Visual	Casi nunca	Raramente	A veces	Frecuentemente	Casi siempre
	12	Me gusta ir a lugares donde tienen luces brillantes y muchos colores.					
	13	Me gusta vestirme con ropa colorida.					
	14	Me siento frustrado cuando trato de encontrar algo en un cajón lleno de cosas o en un salón desordenado.					
	15	No encuentro la calle, el edificio o una habitación dentro de una casa cuando voy a lugares nuevos.					
	16	Me molestan las imágenes visuales que se mueven rápido o son inestables en el cine o la TV.					
	17	No me doy cuenta cuando la gente entra en un lugar/sala.					
	18	Me molesta cuando veo mucha gente alrededor mío (por ejemplo; en un centro comercial lleno, desfile, carnaval).					
	19	Limito las distracciones cuando trabajo (por ejemplo, cierro la puerta o apago la TV).					

Comentarios:

ITEM		D. Procesamiento Tacto	Casi nunca	Raramente	A veces	Frecuentemente	Casi siempre
	20	No me agradan ciertas texturas de la comida (por ejemplo; piel del garbanzo, jugo de manzana, queso, patatas de bolsa).					
	21	Me alejo cuando otros se ponen muy cerca de mí.					
	22	No noto cuando mis manos o cara están sucias.					
	23	Tengo rasguños o moretones, pero no recuerdo cómo me los hice.					
	24	Evito hacer filas o pararme cerca de otras personas porque no me gusta estar muy cerca de otros.					
	25	No noto cuando alguien toca mi brazo o mi espalda.					

Comentarios:

ITEM		E. Nivel de Actividad	Casi nunca	Raramente	A veces	Frecuentemente	Casi siempre
—	26	Me tomo más tiempo que a otras personas para despertarme por las mañanas.					
~	27	Hago cosas por impulso o en el momento (en otras palabras, hago cosas sin tener un plan en mente).					
—	28	Encuentro tiempo para alejarme de mi ajetreada vida y estar solo.					
—	29	Parezco más lento que otros cuando intento hacer una actividad o una tarea.					
—	30	Me mantengo alejado de las aglomeraciones.					
~	31	Encuentro actividades para realizar en público (ej. Música; deportes, actuación, hablar en público y responder preguntas en clase).					
⊖	32	Me resulta difícil concentrarme durante todo el tiempo cuando estoy sentado en una clase o reunido.					

Comentarios:

ITEM		F. Procesamiento Auditivo	Casi nunca	Raramente	A veces	Frecuentemente	Casi siempre
~	33	Tarareo, silbo, canto o hago ruidos.					
⊖	34	Me sorprende fácilmente con ruidos inesperados o fuertes (ej. aspiradora, ladridos de perros, llamadas telefónicas).					
—	35	Tengo problemas para seguir lo que la gente está diciendo cuando hablan rápido o son temas poco conocidos para mí.					
—	36	Huyo de espacios donde otros están viendo la TV, o les pido que bajen el volumen.					
⊖	37	Me distraigo si hay mucho ruido alrededor mío.					
—	38	Uso estrategias para disminuir el sonido (ej. cierro la puerta, tapo mis oídos, uso tapones para oídos).					
—	39	Me mantengo alejado de lugares ruidosos.					
~	40	Me gusta ir a eventos con mucha música.					
—	41	Tengo que pedir a las personas que repitan las cosas.					
⊖	42	Encuentro difícil trabajar con ruido de fondo (ej. ventilador, radio).					

Comentarios:



HOJA DE PUNTAJE

CUADRANTE

Instrucciones: Transfiera del *Auto Cuestionario* el puntaje en bruto del ítem que corresponda a cada ítem del listado (refiérase al Manual del Usuario para seguir las instrucciones de cómo obtener el puntaje bruto). Sume el Puntaje en Bruto de la columna para obtener el Puntaje Bruto Total para cada cuadrante.

 Cuadrante 1		 Cuadrante 2		 Cuadrante 3		 Cuadrante 4	
BAJO REGISTRO		BÚSQUEDA SENSORIAL		SENSIBILIDAD SENSORIAL		EVITACIÓN SENSORIAL	
ITEM	Puntaje Bruto	ITEM	Puntaje Bruto	ITEM	Puntaje Bruto	ITEM	Puntaje Bruto
3		1		5		7	
8		2		11		19	
10		4		14		21	
15		6		16		24	
17		9		18		28	
22		12		20		30	
23		13		32		36	
25		27		34		38	
26		31		37		39	
29		33		42			
35		40					
41							
Puntaje Total Bruto Cuadrante		Puntaje Total Bruto Cuadrante		Puntaje Total Bruto Cuadrante		Puntaje Total Bruto Cuadrante	

REFERENCIA PUNTUACIÓN	
1	Casi Nunca
2	Raramente
3	A Veces
4	Frecuentemente
5	Casi Siempre

REFERENCIA ICONO	
	BAJO REGISTRO
	BÚSQUEDA SENSORIAL
	SENSIBILIDAD SENSORIAL
	EVITACIÓN SENSORIAL

RESUMEN DE CUADRANTES

Instrucciones: Elija el Gráfico de Resumen de Cuadrante más apropiado y transfiera en el Total del Puntaje Bruto del Cuadrante de la página anterior al casillero que corresponda del Gráfico de Resumen de Cuadrante. Marcar estos totales con una X en la columna de clasificación apropiada (Mucho menos que la mayoría de las personas, Menos que la mayoría de las personas, etc.).*

GRÁFICO DE RESUMEN DEL CUADRANTE PARA EDADES DE 18 – 64 AÑOS.

Cuadrante	Puntaje Bruto Total Cuadrante	Mucho menos que la mayoría de las personas	Menos que la mayoría de las personas	Similar a la mayoría de las personas	Más que la mayoría de las personas	Mucho más que la mayoría de las personas
		— —	—	=	+	++
1. Bajo Registro	/60	12 ----- 19	20 ----- 23	24 ----- 25	25 ----- 30	31 ----- 60
2. Búsqueda Sensorial	/55	11 ----- 30	31 ----- 34	35 ----- 37	38 ----- 41	42 ----- 55
3. Sensibilidad Sensorial	/50	10 ----- 21	22 ----- 25	26 ----- 29	30 ----- 34	35 ----- 50
4. Evitación Sensorial	/45	9 ----- 19	20 ----- 22	23 ----- 25	26 ----- 29	30 ----- 45

*Clasificación basada en el desempeño de personas sin discapacidad (n = 806)

