



Universidad
Rey Juan Carlos

TESIS DOCTORAL

**REVISIÓN SISTEMÁTICA DE LOS CUESTIONARIOS
AUTOCUMPLIMENTADOS ADAPTADOS AL ESPAÑOL PARA LA
VALORACIÓN FUNCIONAL DE PACIENTES CON PATOLOGÍA DE
HOMBRO, RODILLA Y CADERA.**

Autor:

Sara Gómez Valero

Director:

Dr. Juan Carlos Miangolarra Page

Dra. Isabel María Alguacil Diego

Programa de Doctorado en Ciencias de la Salud

Escuela Internacional de Doctorado

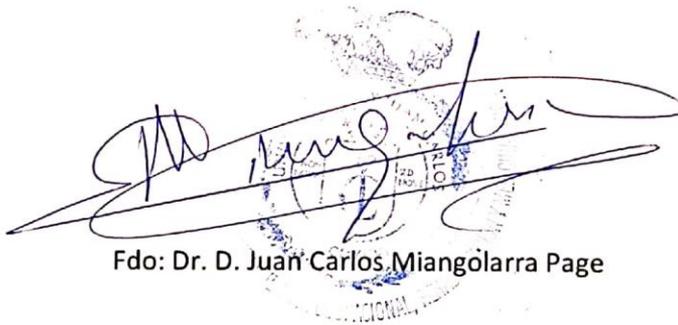
2019

D. Juan Carlos Miangolarra Page, Profesor Titular del Departamento de Fisioterapia, Terapia Ocupacional, Rehabilitación y Medicina Física de la Facultad de Ciencias de la Salud (Universidad Rey Juan Carlos) y Jefe del Servicio de Rehabilitación y Medicina Física del Hospital Universitario de Fuenlabrada

CERTIFICA:

Que el proyecto de tesis doctoral titulado **“Revisión sistemática de los cuestionarios autocumplimentados adaptados al español para la valoración funcional de pacientes con patología de hombro, rodilla y cadera”**, ha sido realizado por Dña. Sara Gómez Valero, bajo mi supervisión y dirección, cumpliendo todos los requisitos científicos y formales para optar al grado de Doctor.

Y para que así conste a todos los efectos, se firma el presente certificado en Madrid a 04 de febrero de 2019.



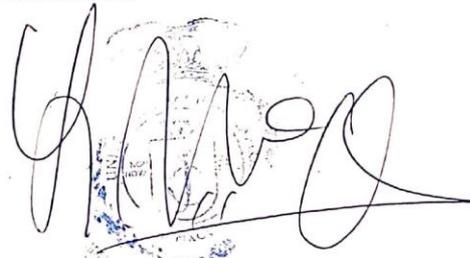
Fdo: Dr. D. Juan Carlos Miangolarra Page

DÑA. ISABEL M^a ALGUACIL DIEGO, Profesora Contratada Doctora del Departamento de Fisioterapia, Terapia Ocupacional, Rehabilitación y Medicina Física, de la Facultad de Ciencias de la Salud

CERTIFICA:

Que el proyecto de tesis doctoral titulado "REVISIÓN SISTEMÁTICA DE LOS CUESTIONARIOS AUTOCUMPLIMENTADOS ADAPTADOS AL ESPAÑOL PARA LA VALORACIÓN FUNCIONAL DE PACIENTES CON PATOLOGÍA DE HOMBRO, RODILLA Y CADERA", ha sido realizado bajo mi dirección por Dña. Sara Gómez Valero y reúne todos los requisitos científicos y formales para ser presentado y defendido ante el tribunal correspondiente.

Y para que así conste a todos los efectos, se firma el presente certificado en Madrid a 7 de marzo de dos mil diecinueve.



Fdo: Dra. Dña. Isabel M^a Alguacil Diego.

A mi marido Jesús y a mi hijo por ser el motor de mi vida.

AGRADECIMIENTOS

Este trabajo de Tesis, realizado en la Universidad Rey Juan Carlos de Madrid, es un esfuerzo en el que, directa o indirectamente, han participado muchas personas opinando, corrigiendo, teniendo paciencia, dando ánimo, acompañando en los momentos de crisis y en los momentos de felicidad, lo que me ha permitido aprovechar su competencia y experiencia que deseo agradecer en este apartado.

En primer lugar, a mis directores de tesis, Dr. Juan Carlos Miangolarra Page y Dra. Isabela Alguacil Diego, mi más sincero agradecimiento por haber confiado en mí para este trabajo, por su disponibilidad, por su paciencia ante mi inconsistencia, por su valiosa dirección y su capacidad para guiar mis ideas y apoyo para llegar a la conclusión del mismo.

Al servicio de Medicina Física y Rehabilitación del Hospital Universitario Fundación Alcorcón por apoyarme y ayudarme en mis años de formación como residente y, en especial, a los Dres. Mariano Tomas Flórez García y Fernando García Pérez, por sus conocimientos y esfuerzo, siendo piezas claves e indispensables en este trabajo.

A todos mis compañeros y amigos, se merecen muchas y buenas palabras, ya que con ellos he compartido horas de trabajo, de convivencia y buenos ratos, lo cual no tiene precio. Por todo el tiempo que me han dado, por las conversaciones de las que tanto provecho he sacado, por su respaldo y amistad. Gracias por los buenos y malos momentos, por aguantarme y por escucharme. Siempre estuvieron ahí en lo bueno y en lo malo, en ocasiones con más confianza en el trabajo que yo misma. ¿Quién puede pedir más? A todos ellos, gracias.

Y, por supuesto, el agradecimiento más profundo y sentido va para mi familia. Sin su apoyo, colaboración e inspiración habría sido imposible llevar a cabo este reto. En especial, a mis padres, Raquel y Fernando, por su ejemplo de cariño incondicional y honestidad; a mi hermano Fernando, por su ejemplo de valentía y su capacidad de superación.

Las palabras nunca serán suficientes para testimoniar mi aprecio y mi agradecimiento.

En suma, a todos ellos, mi mayor reconocimiento y gratitud.

RESUMEN

Introducción. Dentro de los sistemas de evaluación de los resultados de los tratamientos un aspecto esencial es que el propio paciente, mediante cuestionarios, valore desde su perspectiva los resultados alcanzados. Aunque existen numerosos cuestionarios para medir la limitación funcional percibida por el paciente para las patologías de hombro, rodilla y cadera, la mayoría se han diseñado en países anglosajones. Para utilizarlos en otros idiomas, países y/o culturas se necesita, primero, realizar una adaptación transcultural y, después, comprobar que la escala mantiene la fiabilidad y la validez de la versión original.

En el mundo sanitario, desde hace años se emplean técnicas de trabajo en grupo para disminuir los posibles sesgos en los estudios.

Las recomendaciones basadas en el criterio de los expertos siguen siendo una valiosa ayuda para tomar decisiones en la práctica rutinaria.

Objetivo. Realizar una revisión sistemática sobre los cuestionarios autocumplimentados de hombro, rodilla y cadera adaptados al español, analizando tanto la calidad metodológica del proceso de adaptación transcultural como las propiedades psicométricas de la versión española. En base a ello, seleccionar el cuestionario de hombro más indicado para su uso en la práctica clínica.

Material y método. En una primera fase, se realizó una búsqueda en las principales bases de datos biomédicas para localizar las de escalas de valoración funcional de hombro, rodilla y cadera adaptadas al español. Dos de los autores revisaron los artículos seleccionados y valoraron tanto si el proceso de adaptación del cuestionario había seguido las recomendaciones internacionales como si se evaluaron apropiadamente sus propiedades psicométricas.

En una segunda fase, se procedió a localizar y analizar las escalas adaptadas para la articulación del hombro. Mediante consenso de expertos se seleccionó el cuestionario

de hombro considerado como el más idóneo para la evaluación del paciente en la práctica clínica y se incorporó al sistema informático del hospital para facilitar su uso.

Resultados. La estrategia de búsqueda identificó un total de 21 cuestionarios. Dos aplicables a cualquier tipo de afección del miembro superior: *Disabilities of the Arm, Shoulder and Hand Questionnaire* (DASH) y *Upper Limb Functional Index* (ULFI). Cuatro cuestionarios generales de hombro: *Simple Shoulder Test* (SST), *Shoulder Pain and Disability Index* (SPADI), *Oxford Shoulder Score* (OSS) y *Shoulder Disability Questionnaire* (SDQ). Y tres para situaciones específicas: *Western Ontario Rotator Cuff Index* (WORC) para sujetos con patología manguito rotador, *Western Ontario Shoulder Instability Index* (WOSI) para pacientes con inestabilidad de hombro y *Wheelchair Users Shoulder Pain Index* (WUSPI) para usuarios de sillas de ruedas. Tres eran aplicables para patologías del miembro inferior. Dos para cualquier tipo de afección: *Lower Limb Functional Index* (LLFI), *Lower Extremity Functional Scale* (LEFS) y uno para artrosis: *Arthrose des Membres Inférieurs et Qualité de vie* (AMICAL). Tres para patologías de rodilla y cadera: *Western Ontario and McMaster Universities Osteoarthritis index* (WOMAC), *Osteoarthritis Knee and Hip Quality of Life* (OAKHQOL) y *Hip and Knee Questionnaire* (HKQ). Cuatro para patologías de rodilla. Dos generales: *Knee Injury and Osteoarthritis Outcome Score* (KOOS) y *Knee Society Clinical Rating System* (KSS) y dos específicas: *Victorian Institute of Sport Assessment questionnaire* (VISA-P) para la tendinopatía rotuliana y *Kujala* para pacientes con dolor femoropatelar. Y dos exclusivos de cadera: *Hip Outcome Score* (HOS) e *International Hip Outcome Tool-33* (iHOT-33).

Siete de ellos publicados en los últimos dos años (SST, SPADI (X2), OSS, WORC, WOSI, iHOT-33), y 6 en 2014 (LLFI; LEFS; AMICAL; HKQ; HOS; KUJALA). De todos se dispone, también, de adaptaciones transculturales a otros idiomas.

Aunque todavía no hay un consenso internacional sobre el modo correcto de realizar la adaptación transcultural, sí parece haber acuerdo en que la traducción inversa o retrotraducción y la prueba piloto (pretest) resultan esenciales.

Los principales criterios reconocidos internacionalmente referentes al proceso de adaptación al español se siguieron con el máximo rigor para 13 cuestionarios (DASH x3, SPADI X2, OSS, WORC, WOSI, AMICAL, KOOS, KSS, OAKHQOL, VISA-P, KUJALA, HOS, IHOT-33), pero también los restantes se pueden considerar metodológicamente válidos, aunque el proceso fue algo menos riguroso para los cuestionarios ULFI, SST, SDQ, WUSPI, HKQ y LLFI .

En las diferentes naciones o poblaciones hispanohablantes pueden existir palabras de uso exclusivo o con significado diferente a otros países o grupos culturales. No obstante, en general, no hay diferencias culturales tan marcadas que impidan el uso de estos cuestionarios en estos países.

Las propiedades psicométricas analizadas fueron aceptables y similares a la versión original y a otras versiones adaptadas a otros idiomas.

Al finalizar la reunión de consenso de expertos se decidió que la escala de hombro SPADI sería la escala que se integraría en el sistema informático del hospital para evaluar a los pacientes.

ABSTRACT

Introduction. An essential aspect of treatment-outcome evaluation systems is that the patients themselves use questionnaires to assess the outcomes achieved from their own perspective. There are numerous questionnaires to measure functional limitation perceived by patients suffering joint pathologies in shoulder, knee and hip but most of these have been developed in English-speaking countries. In order to be used in other countries, languages and/or cultures, these questionnaires need a transcultural adaptation and maintenance of reliability and validity must be checked.

In healthcare environment, group work techniques have been used regularly to reduce, as much as possible, biases in studies.

Recommendations based on expert judgment continue to be a valuable aid in making decisions in routine practice.

Objective. To conduct a systematic review, carrying out a comprehensive literature search, of self-administered shoulder, knee and hip-disability functional assessment questionnaires adapted to Spanish, analyzing the quality of the transcultural adaptation procedure and the clinimetric properties of the new version. Also, select the more indicated shoulder questionnaire for its use in clinical practice.

Material and methods. In a first phase, we made a search in the main biomedical databases to locate all shoulder, knee and hip function assessment scales adapted to Spanish. Two authors reviewed papers selected and considered either the process of adaptation of the questionnaire had followed international recommendations and if its psychometric properties had been appropriately assessed.

In a second phase, we proceeded to locate and analyze the shoulder scales already adapted. By consensus of experts, the shoulder questionnaire considered most suitable for the evaluation of the patient in clinical practice was selected and incorporated into the computer system of the hospital.

Results. The search strategy identified twenty one relevant publications. Two applicable to upper limb: *Disabilities of the Arm, Shoulder and Hand Questionnaire* (DASH) and *Upper Limb Functional Index* (ULFI). Four general shoulder questionnaires: *Simple Shoulder Test* (SST), *Shoulder Pain and Disability Index* (SPADI), *Oxford Shoulder Score* (OSS) and *Shoulder Disability Questionnaire* (SDQ), and three for specific situations: *Western Ontario Rotator Cuff Index* (WORC) for subjects with rotator cuff pathology, *Western Ontario Shoulder Instability Index* (WOSI) for patients with shoulder instability and *Wheelchair Users Shoulder Pain Index* (WUSPI) for wheelchair users. Three were applicable for lower limb pathologies, two for any type of condition: *Lower Limb Functional Index* (LLFI), *Lower Extremity Functional Scale* (LEFS) and one for osteoarthritis: *Arthrose of Membres Inférieurs et Qualité de vie* (AMICAL). Three for knee and hip pathologies: *Western Ontario and McMaster Universities Osteoarthritis index* (WOMAC), *Osteoarthritis Knee and Hip Quality of Life* (OAKHQOL) and *Hip and Knee Questionnaire* (HKQ). Four for knee pathologies, two general: *Knee Injury and Osteoarthritis Outcome Score* (KOOS) and *Knee Society Clinical Rating System* (KSS) and two specific: *Victorian Institute of Sport Assessment questionnaire* (VISA-P) for patellar tendinopathy and *Kujala* score for patients with femoropatellar pain. And two exclusive for hip: *Hip Outcome Score* (HOS) and *International Hip Outcome Tool-33* (iHOT-33).

Seven of them were published in the last two years (SST, SPADI (X2), OSS, WORC, WOSI, iHOT-33) and 6 in 2014 (LLFI; LEFS; AMICAL; HKQ; HOS; KUJALA). All of them also have transcultural adaptations to other languages.

There is no clear international consensus about the correct way to carry out cross-cultural adaptation. However, there seems to be agreement that reverse translation or backtranslation and the pilot test (pretest) are essential.

The internationally recognized criteria concerning the adaptation process to Spanish were followed with the maximum rigor for 13 questionnaires (DASH x3, SPADI X2, OSS, WORC, WOSI, AMICAL, KOOS, KSS, OAKHQOL, VISA-P, KUJALA, HOS, IHOT -33) but also the others can be considered methodologically valid, although the process was less rigorous for ULFI, SST, SDQ, WUSPI, HKQ and LLFI questionnaires. In different Spanish-speaking nations or populations there may be words of exclusive use or with a different meaning to other countries or cultural groups. However, in general, there are no marked cultural differences which prevent the use of these questionnaires among them.

The psychometric properties analyzed were acceptable and similar to both the original and the adapted version to other languages.

At the end of the expert consensus meeting, it was decided that SPADI shoulder scale should be the scale to be integrated into the hospital's computer system to assess patients.

ÍNDICE

1. GLOSARIO.....	29
2. RELACIÓN DE TABLAS.....	33
3. RELACIÓN DE FIGURAS.....	37
4. INTRODUCCIÓN.....	41
5. JUSTIFICACIÓN DEL ESTUDIO	67
6. HIPÓTESIS.....	71
7. OBJETIVOS.....	75
8. MATERIAL Y MÉTODOS.....	79
8.1. PRIMERA FASE.....	81
8.1.1. Estrategia de búsqueda.....	81
8.1.2. Criterios de selección.....	83
8.1.3. Análisis de la información.....	87
8.1.3.1. Características de los participantes en el estudio.....	87
8.1.3.2. Evaluación de la metodología utilizada para efectuar la adaptación transcultural.....	87
8.1.3.3. Evaluación de las propiedades psicométricas de la versión española	90
8.1.3.4. Evaluación de la aplicabilidad directa de las escalas adaptadas un país hispanohablante para ser aplicada en otro	91
8.2. SEGUNDA FASE.....	91
8.2.1. Selección de expertos.....	92
8.2.2. Criterios de selección de expertos.....	93
8.2.3. Convocatoria.....	94
8.2.4. Reunión de expertos.....	94
8.2.5. Participantes.....	95

8.2.6. Desarrollo de la reunión	96
9. RESULTADOS	101
9.1. PRIMERA FASE.....	103
9.1.1. Hombro.....	103
9.1.2. Rodilla.....	113
9.1.3. Cadera.....	120
9.2. SEGUNDA FASE.....	127
10. DISCUSIÓN.....	129
10.1. Escalas de valoración	131
10.1.1. Escalas de hombro	134
10.1.2. Escalas de rodilla y cadera	136
10.2. Panel de expertos.....	139
11. LIMITACIONES Y PUNTOS FUERTES DEL ESTUDIO.....	143
12. CONCLUSIONES.....	147
13. BIBLIOGRAFÍA.....	151
14. ANEXOS.....	175
14.1. Anexo 1. E-mail enviado a los participantes del grupo de expertos	177
14.2. Anexo 2. Artículo de los cuestionarios autocumplimentados adaptados al español para la valoración funcional de pacientes con afecciones de hombro	179
14.3. Anexo 3. Artículo de los cuestionarios autocumplimentados adaptados al español para la valoración funcional de pacientes con afecciones de rodilla	193
14.4. Anexo 4. Artículo sobre los cuestionarios autocumplimentados adaptados al español para la valoración funcional de pacientes con patología de cadera	201

1. GLOSARIO

AMICAL: Arthrose des Membres Inférieurs et Qualité de vie

AVD: Actividades de la vida diaria

CIDDM: Clasificación Internacional de Deficiencias, Discapacidades y Minusvalías

CIF: Clasificación Internacional del Funcionamiento, la Discapacidad y la Salud

CMD: Cambio mínimo detectable

DASH: Disabilities of the Arm, Shoulder and Hand Questionnaire

EMM: Error mínimo estandarizado

HKQ: Hip and Knee questionnaire

HOS: Hip Outcome Score

iHOT-33: International Hip Outcome Tool-33

ICC: Índice de correlación intraclase

KOOS: Knee Injury and Osteoarthritis Outcome Score

KSS: Knee Society Clinical Rating System

LEFS: Lower Extremity Functional Scale

LLFI: Lower Limb Functional Index

MFR: Medicina Física y Rehabilitación

OAKHQOL: Osteoarthritis Knee and Hip Quality of Life questionnaire

OMS: Organización Mundial de la Salud

OMS: Organización Mundial de la Salud

PIB: Producto interior bruto

OSS: Oxford Shoulder Score

RME: Respuesta media estandarizada

SPADI: Shoulder Pain and Disability Index

SDQ: Shoulder Disability Questionnaire

SST: Simple Shoulder Test

TE: Tamaño del efecto

ULFI: Upper Limb Functional Index

VISA-P: Victorian Institute of Sport Assessment questionnaire

WOMAC: Western Ontario and McMaster Universities Osteoarthritis index

WORC: Western Ontario Rotator Cuff Index

WOSI: Western Ontario Shoulder Instability Index

WUSPI: Wheelchair Users Shoulder Pain Index

2. RELACIÓN DE TABLAS

Tabla 1. Ejemplo simplificado del core set en su idioma original para pacientes con patología musculoesquelética	50
Tabla 2. Comparativa de los principales métodos de consenso. Ventajas e inconvenientes. Reproducida de Peiró S, Protella E. 2003.....	66
Tabla 3. Participantes en el panel de expertos.....	96
Tabla 4. Características demográficas y clínicas de la población de los estudios de adaptación transcultural.....	106
Tabla 5. Evaluación de la metodología empleada para la adaptación transcultural de los cuestionarios	108
Tabla 6. Principales propiedades psicométricas analizadas en los cuestionarios adaptados	109-110
Tabla 7. Características demográficas y clínicas de la población de los estudios de adaptación transcultural.....	115
Tabla 8. Evaluación de la metodología empleada para la adaptación transcultural de los cuestionarios	116

Tabla 9. Principales propiedades psicométricas analizadas en los cuestionarios adaptados	118-119
Tabla 10. Características demográficas y clínicas de la población de los estudios de adaptación transcultural.....	121-122
Tabla 11. Evaluación de la metodología empleada para la adaptación transcultural de los cuestionarios	123
Tabla 12. Principales propiedades psicométricas analizadas en los cuestionarios adaptados	125-126

3. RELACIÓN DE FIGURAS

Figura 1. Diagrama del ejemplo de aplicación del modelo CIF en enfermedades músculo-esqueléticas	48
Figura 2. Diagrama de flujo que muestra la selección de los estudios incluidos para patologías de hombro	84
Figura 3. Diagrama de flujo que muestra la selección de los estudios incluidos para patologías de rodilla	85
Figura 4. Diagrama de flujo que muestra la selección de los estudios incluidos para patologías de cadera	86
Figura 5. Síntesis de la metodología para la traducción y adaptación cultural de las escalas adaptado de Beaton et al.....	89
Figura 6.- Desarrollo de la reunión del panel de expertos.....	99

4. INTRODUCCIÓN

El dolor físico es una de las dolencias más comunes de la población general y la causa más frecuente de consulta en Atención Primaria¹.

Hay que diferenciar el dolor agudo, que siempre es limitado en el tiempo, del dolor crónico, que se prolonga durante más de tres meses y a menudo desencadena un conjunto de cambios físicos y psicosociales, como dependencia, depresión del sistema inmunológico y mayor vulnerabilidad a enfermar, alteraciones del sueño y del apetito, bajo rendimiento laboral, además de miedo, angustia y depresión².

Una de cada cinco personas sufre dolor crónico en Europa. España es el país europeo en donde más tiempo tardan en controlar su dolor las personas y dos de cada tres no lo consiguen en ningún momento³. Debemos tratar cuanto antes el dolor, básicamente por dos motivos: para mitigar el sufrimiento del paciente, y para evitar o disminuir los importantes cambios que provoca y así impedir llegar a una situación de mayor gravedad.

Aunque el coste del dolor crónico es difícil de calcular, ya que no se dispone de datos globales, en España se estima que el coste total que ocasiona el dolor crónico sería de 16.000 millones de euros anuales, lo que supone el 2,5% del producto interior bruto (PIB). Y de estos, la osteoartritis de rodilla y cadera suponen un coste anual de 4.738 millones de euros (equivalente al 0,5% del PIB)⁴.

Las patologías musculo-esqueléticas son comunes en la comunidad adulta, estimándose un aumento de su prevalencia todavía mayor debido al crecimiento de la población, el envejecimiento y los estilos de vida sedentarios, especialmente en los

países desarrollados⁵. Son una de las principales causas de morbilidad, limitación de la actividad, incapacidad funcional y uso de los servicios de salud⁶.

El dolor de hombro es una de las afecciones musculoesqueléticas más comunes observadas en atención primaria⁷, después de la columna lumbar y cervical, con una prevalencia en España del 17,12%⁸, seguida de las afecciones de rodilla (10,2%) y de cadera (4%)⁹.

El “hombro doloroso” no es un diagnóstico específico, pues engloba diferentes diagnósticos de límites poco precisos, y que, además se solapan entre sí¹⁰⁻¹².

Dentro de las patologías de hombro, la causa más habitual de dolor es la patología del manguito rotador, que es una inflamación de un grupo tendinoso que rodea la cápsula articular de la articulación del hombro¹³. En cuanto a su etiología, la inflamación de los tendones de los músculos del hombro, especialmente del manguito de los rotadores, puede presentarse por la repetición de movimientos, sobre todo las maniobras de abducción. Esta inflamación viene causada por la estrechez por donde trascurren los tendones, que favorece el rozamiento que termina produciendo la inflamación del tendón.

En cuanto a la incidencia se presenta entre el 0,9-2,5% (en función de los grupos de edad) y la prevalencia entre 6,9-66,7%, en función del periodo de estudio y los criterios de definición de hombro doloroso¹⁴. En España se han descrito cifras de prevalencia de 78 por 1.000 habitantes¹⁵ y los estudios de revisión relatan variaciones en prevalencia entre 70-200 por 1.000 adultos¹⁶.

La patología del hombro es la segunda causa de incapacidades laborales médicas largas o prolongadas después de la lumbalgia¹⁷. De hecho, constituye una de las 20 primeras patologías en alcanzar los 12 meses de baja (incapacidad temporal) e incluso entre los procesos que causan prolongación de la misma, es decir que prolongan la situación de incapacidad laboral hasta los 545 días, y su incidencia es así mismo relevante entre los 20 procesos que sobrepasan los 6 meses, por ello pudiéramos considerar que la patología con hombro doloroso es causa de incapacidad médica prolongada. El problema del absentismo por hombro doloroso reside en que pesar de la benignidad presumible del proceso, es decir resolución esperada del proceso no más allá de 4-6 semanas, un 50% sigue presentando clínica después de 18 meses. A esto hay que añadir la alta recurrencia: en torno al 30%¹⁸.

Los factores de riesgo y la sintomatología que caracterizan al dolor en el hombro son variables según la causa que lo determina, siendo común el dolor cuando la persona trata de mover el brazo por encima de la cabeza, detrás de la espalda o hacia el frente, una postura antiálgica (hombro en rotación interna, aducción y mano doblada sobre abdomen) y una clara limitación funcional. Los objetivos generales del tratamiento incluyen erradicar los síntomas, evitar la cronicidad, mejorar el movimiento y la funcionalidad de su extremidad superior¹⁰⁻¹².

El dolor de rodilla en adultos es, también, una causa frecuente de consulta en atención primaria¹⁹: alrededor del 80% de las personas menores de 40 años lo han presentado, de forma incapacitante, al menos una vez en su vida, siendo las causas más frecuentes las tendinitis, las bursitis, y las meniscopatías²⁰.

La rodilla es la principal gran articulación que se ve afectada en la artrosis, siendo el riesgo de sufrir incapacidad por gonartrosis tan grande como el secundario a enfermedades cardiológicas y mayor que por cualquier otra patología médica en la ancianidad²¹. Un informe reciente de la Organización Mundial de la Salud (OMS) sobre la carga global de la enfermedad indica que la artrosis de rodilla está próxima a ser la cuarta causa más importante de discapacidad en mujeres y la octava en varones, siendo los costes anuales atribuidos a tal patología inmensos²².

La cadera es la otra articulación más frecuentemente afectada por artrosis, después de la rodilla, y una de las que más incapacita²³.

La causa del dolor de cadera es casi siempre multifactorial. Una serie de factores de riesgo producen inestabilidad, desviación, aumento de la carga articular, microtraumatismos y daño estructural. La articulación responde con inflamación subcondral y sinovial e hipertrofia ósea²⁴.

El dolor en cadera es un síntoma inespecífico y mal definido, que frecuentemente es motivo de consulta e implica un amplio diagnóstico diferencial. El lograr un enfoque diagnóstico y terapéutico acertado suele facilitarse si nos apoyamos en la historia clínica y la exploración física, disminuyendo la necesidad de pruebas de imagen o visitas al especialista, lo que permite acelerar el proceso curativo. El diagnóstico rápido y la instauración de las alternativas terapéuticas adecuadas disminuyen el dolor, mejoran la funcionalidad y reducen el riesgo de incapacidad a largo plazo²⁵, por lo que estas medidas están en la base de la calidad de la atención prestada a estas patologías.

En el entorno sanitario actual, es cada vez más importante poder cuantificar los cambios asociados con una intervención dada. Esto se puede lograr usando una o más

medidas de resultado apropiadas. Sin embargo, la selección e integración de medidas de resultado dentro de la práctica clínica requiere una meticulosa consideración²⁶.

En el mes de mayo de 2000, la 54ª Asamblea de la OMS aprobó en su Resolución WHA/21 la “Clasificación Internacional del Funcionamiento, la Discapacidad y la Salud” (CIF), una nueva clasificación que viene a sustituir a la Clasificación Internacional de Deficiencias, Discapacidades y Minusvalías (CIDDM), que venía utilizándose desde 1980, y en la que, a diferencia de la actual, los parámetros se establecieron en base a las consecuencias que la enfermedad producía en la persona. Esta nueva clasificación aporta un enfoque mayoritariamente centrado en los componentes de salud del ser humano y de este modo proporciona una visión más positiva.

Se ha producido por tanto un giro fundamental, pasándose del binomio sujeto-enfermedad al de sujeto-salud²⁷. Esto implica la selección de medidas de resultado basadas en la necesidad, idoneidad y viabilidad y un uso cuidadoso en la práctica clínica habitual, incluidos la recopilación de datos, el análisis y la reevaluación del proceso.

Describiremos este cambio, centrándonos en las patologías musculoesqueléticas²⁶.

La CIF integra los modelos biomédico y social de la discapacidad y asume el modelo biopsicosocial como marco conceptual. Además, en la medida en que recoge factores ambientales y personales en su modelo, permite describir cómo éstos pueden ser un elemento clave para comprender la discapacidad y el funcionamiento de los

individuos y las poblaciones, y cómo se pueden adoptar cambios y medidas sociales que permitan reducir su impacto cuando éste es negativo²⁸.

En la figura 1 se presenta un ejemplo de aplicación del modelo CIF para pacientes con enfermedades músculo esqueléticas²⁹.

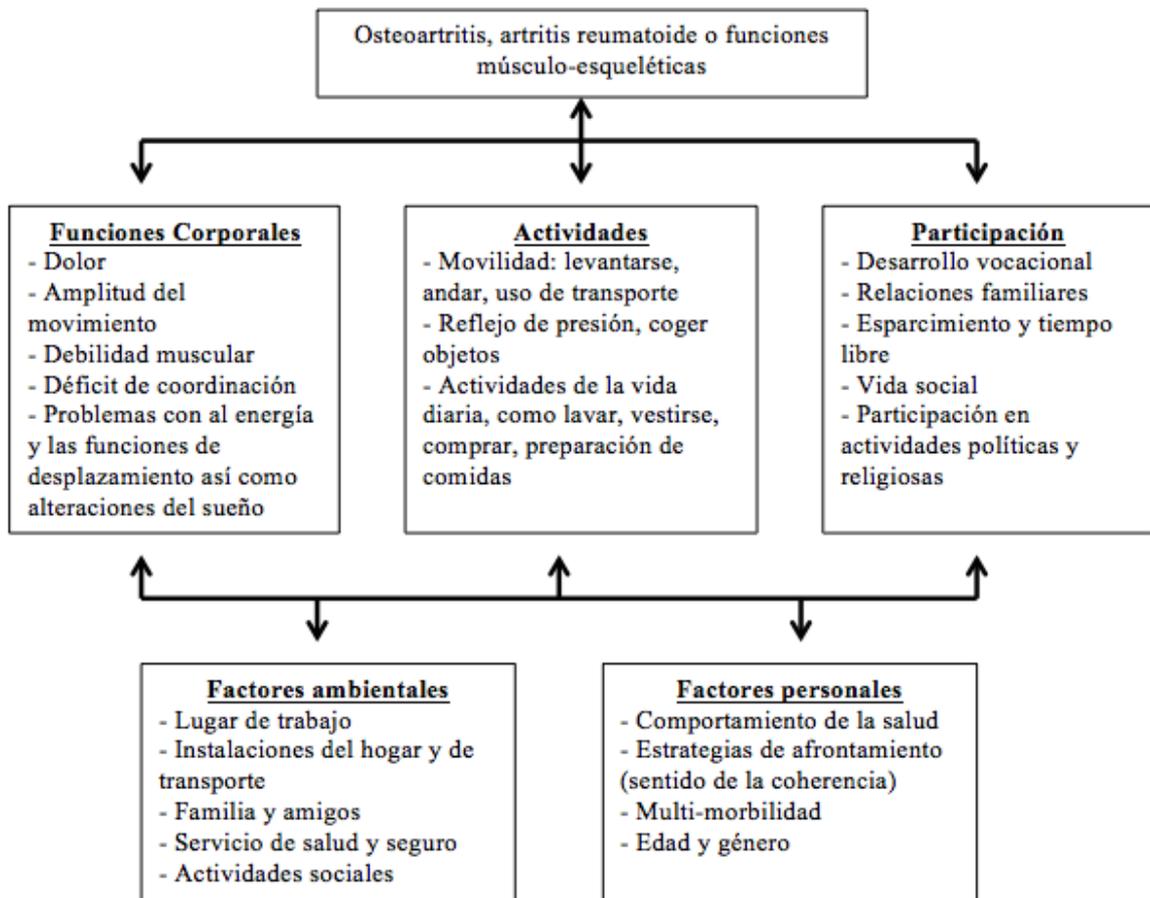


Figura 1. Ejemplo de aplicación del modelo CIF en enfermedades músculo-esqueléticas²⁹.

La CIF propone la existencia de categorías que incluyen las funciones y estructuras corporales, así como las actividades, participación y factores ambientales.

El sistema que utiliza es alfanumérico: los códigos utilizados comienzan con una letra que corresponde a los diferentes ítems en su idioma original (inglés):

- b: Escala de Funciones corporales (body);
- s: Escala de Estructuras corporales (structure);
- d: Escala de actividad y participación (disability);
- e: Escala de factores ambientales (environment).

Los conjuntos básicos o *core set* son listas reducidas que contienen las categorías de la CIF más relevantes y significativas en la descripción de todo el espectro de problemas y alteraciones en el funcionamiento de los individuos con una enfermedad específica.

Se evalúan no solo la afectación de las funciones y estructuras (dolor, rango de movimiento, debilidad, déficit de coordinación), sino también la limitación en actividades (movilidad, manejo y agarre de objetos, actividades de la vida diaria) y las restricciones en participación (desempeño vocacional, relaciones familiares, recreo y ocio, vida social), modificados por los factores contextuales (factores ambientales y personales, comportamiento de salud, estrategias de afrontamiento, co-morbilidad, edad y sexo), así como la calidad de vida²⁸. Puede contribuir de manera sustancial en el manejo del paciente con estas patologías para mejorar sus actividades desde una perspectiva personal y social.

En la tabla 1, presentamos un ejemplo simplificado, en el idioma original (inglés), del *core set* para pacientes con patología musculoesquelética^{30,31}.

Tabla 1. Ejemplo simplificado del core set en su idioma original para pacientes con patología musculoesquelética.

CODE	ICF CATEGORIES	INFLUENCE/PROBLEMS	PRM INTERVENTIONS
	BODY FUNCTIONS		
b280	Sensation of pain	Pain in affected joints	Pain management interventions within the scope of PRM
b710	Mobility of joint functions	Impaired mobility and functioning	Exercise and other measures
	BODY STRUCTURES		
S750	Structure of lower extremity	Degenerative changes related OA	Pharmacological and no pharmacological approaches
	ACTIVITIES AND PARTICIPATION		
d450	Walking	Limitations in the ability of walking	Mobility training
	ENVIRONMENTAL FACTORS		
E115	Products and technology for personal use in daily living	May be facilitators for functioning	Assistive devices

La Medicina Física y Rehabilitación (MFR) es una especialidad médica definida por la Sección de Medicina Física y Rehabilitación de la Unión Europea de Especialistas, en su libro blanco³² de la siguiente forma:

“LA MFR es una especialidad médica independiente relativa al funcionamiento físico y cognitivo, a la actividad (incluyendo la conducta), a la participación (incluyendo la calidad de vida) y a la modificación de factores personales y

ambientales. De este modo, es responsable de la prevención, del diagnóstico, de los tratamientos y del manejo y gestión de la rehabilitación de las personas de todo tipo de edad, con procesos discapacitantes y comorbilidad.”

Los especialistas en MFR están involucrados, por tanto, no solo en el diagnóstico y los tratamientos médicos, sino que también deben ocuparse de la mejora de todos los componentes del funcionamiento humano, como se define en la CIF, incluidos los factores personales y ambientales.

Estos especialistas tienen, además, un papel esencial en los equipos interdisciplinarios; su capacitación y experiencia específica les permite diagnosticar y evaluar la gravedad de los problemas de salud, un requisito previo para una intervención segura. Están, por lo tanto, bien capacitados para coordinar los programas de MFR y desarrollar y evaluar nuevas estrategias de gestión. Su amplia formación también implica que pueden tener una visión holística de la atención de un paciente individual³³.

En la literatura científica existen múltiples estudios que han investigado las estrategias usadas en el tratamiento de las patologías musculoesqueléticas que nos ocupan (hombro, rodilla y cadera)³⁴⁻⁴⁰.

Se recomienda que los médicos de MFR se centren en el dolor como objetivo principal de sus intervenciones, en cualquier campo en el que apliquen sus competencias. El tratamiento del dolor debe dirigirse a reducir los síntomas y mejorar el funcionamiento, reducir la discapacidad o recurrencias y mejorar la condición de salud a largo plazo⁴¹.

Hoy en día, se ha generalizado el uso de escalas con el objetivo de conocer la situación basal del paciente, determinar el impacto de la enfermedad actual, transmitir información objetiva, monitorizar los cambios y, en último término, establecer

tratamientos específicos y valorar la respuesta a este aspecto⁴². Es decir, conocer si se han cumplido los resultados o metas que queríamos alcanzar en rehabilitación.

Se han desarrollado numerosos instrumentos para medir los resultados de los tratamientos en afecciones musculoesqueléticas, aplicables tanto en estudios de investigación como en la práctica cotidiana.

Dentro de los sistemas de evaluación de los resultados de los tratamientos un aspecto esencial es que el propio paciente, mediante cuestionarios, valore desde su perspectiva los resultados alcanzados.

Las mediciones de juicios “subjetivos” pueden estimarse cuantitativamente a través de instrumentos denominados “escalas”.

Conocemos como escalas o cuestionarios a aquellas herramientas de evaluación, autoadministradas o administradas por un entrevistador y utilizadas para medir diferencias entre grupos de pacientes en un momento dado o para medir cambios en la salud de pacientes en diferentes momentos.

Es una herramienta que permite asignar valores a ítems de acuerdo a una regla de decisión. Usualmente los valores elegidos son números y las opciones de respuesta de los ítems se definen previamente en tal forma que sean mutuamente excluyentes y exhaustivas⁴³.

Son múltiples las escalas de evaluación funcional desarrolladas en los últimos años. En general todas incluyen diferentes funciones que determinan fundamentalmente

la capacidad de tener cuidado de uno mismo y la movilidad. Las hay genéricas y de aplicación amplia o más específicas orientadas a procesos concretos⁴⁴.

En el campo de la rehabilitación, las escalas permiten hacer valoraciones más objetivas y medir resultados tras determinadas intervenciones.

Un problema con el que nos enfrentamos es la diversidad de escalas que existen y la complejidad que esto representa, no solo para conocerlas y manejarlas, sino para poder comparar los resultados entre distintos estudios. Las escalas nos van a permitir transformar la información clínica en un lenguaje objetivo y universal.

Su uso está cada vez más aceptado y difundido⁴⁵.

Los cuestionarios son un instrumento de evaluación que permite valorar una serie de ítems dando una puntuación global al final de la evaluación, además de estandarizar el procedimiento, cuyo fin es comparar la información, intentando traducir la información clínica a un lenguaje objetivo y universal⁴⁶.

Estos instrumentos se pueden clasificar de diferentes maneras en función de sus características, pero todas ellas deben cumplir las propiedades psicométricas de validez, fiabilidad y sensibilidad al cambio. Estas propiedades se deben evaluar siempre antes de utilizar una escala⁴⁷.

Aunque existen numerosos cuestionarios, tanto genéricos como específicos, para medir la limitación funcional percibida por el paciente, la mayoría han sido elaborados en países anglosajones.

En el caso de que no hayan sido elaborados en el país, región geográfica, cultura o idioma donde vayan a aplicarse es necesario primero, realizar una adecuada adaptación transcultural para que tenga el mismo sentido que la versión original^{48,49}. De ese modo se evitan interpretaciones erróneas por diferencias culturales o en el estilo de vida que pueden limitar o sesgar los resultados obtenidos tras su aplicación. Esas diferencias pueden existir incluso entre países distintos, aunque hablen el mismo idioma.

El proceso de adaptación de una escala de evaluación es clave para asegurar su validez en distintas culturas y para ello debe seguir un método que certifique esa equivalencia, comprobando que conserva las propiedades psicométricas adecuadas para medir aquello para lo que fue diseñada. Su objetivo es conseguir que el instrumento sea equivalente en distintas culturas. Esta calidad metodológica utilizada es imprescindible para asegurar que los resultados que se obtienen en el estudio se puedan interpretar adecuadamente y utilizar en la práctica clínica⁵⁰.

Un proceso de validación deficiente reduce las posibilidades de hacer comparaciones entre poblaciones de distintos países, culturas e idiomas⁵¹.

La mayoría de los estudios realizados sobre la traducción y adaptación de instrumentos utilizan diferentes métodos que incluyen uno o más de los siguientes pasos: traducción del instrumento, retrotraducción, comité evaluador, estudio piloto... Aunque estas técnicas son las recomendadas ninguna de ellas es completa, por lo que se aconseja realizar más de una al mismo tiempo.

Por este motivo, Ramada-Rodilla et al, en 2003, realizaron una revisión bibliográfica exhaustiva para recabar toda la información disponible sobre la

metodología correcta para la adaptación transcultural de los cuestionarios de salud, recomendando dos etapas para el proceso:

- Primera etapa: Traducción y adaptación cultural. Se traduce la escala al idioma de interés partiendo de su versión original y procurando mantener la estructura del cuestionario, para conseguir que el instrumento obtenido mantenga la equivalencia semántica, idiomática y conceptual con el cuestionario original. En la literatura existe un amplio consenso sobre cómo abordar esta primera etapa⁴⁹⁻⁵², recomendándose una secuencia de cinco pasos: 1) Traducción directa del cuestionario original al español (efectuado de forma independiente por al menos dos traductores bilingües cuyo idioma nativo fuera el español); 2) Síntesis de traducciones y solución de posibles discrepancias entre los traductores con algún miembro del equipo de investigación; 3) Traducción inversa o retro-traducción (de la traducción consensuada en español al idioma original, de forma independiente, por al menos dos traductores que no conocieran dicha versión); 4) Revisión por un comité de expertos para desarrollar la versión pre-final asegurando la equivalencia semántica, idiomática, experiencial y conceptual de la escala; y 5) Prueba piloto (pre-test) del cuestionario pre-final con sujetos de habla española con el fin de analizar la viabilidad y aplicabilidad de la escala (se recomienda hacerlo, idealmente, con 30-40 personas, buscando ítems no contestados y posibles problemas de comprensión).
- Segunda etapa: Validación en el idioma destino, para comprobar que conserva todas las propiedades psicométricas de la versión original.

La adaptación transcultural es, por tanto, un proceso más complejo que la simple traducción directa o literal y, además, no garantiza la conservación de las propiedades psicométricas del cuestionario adaptado. Por ello es necesario, siempre, analizar las propiedades psicométricas de la nueva versión para asegurarse de que se trata de una herramienta equivalente a la original y no asumirlo directamente.

Para considerar válida una escala, ésta debe cumplir una serie de características como la sencillez, utilidad y aceptación por parte de los pacientes e investigadores, así como comprobar que la escala mantiene la fiabilidad, la validez y la sensibilidad al cambio de la versión original⁵³.

La fiabilidad es un instrumento que evalúa la exactitud o precisión de los resultados obtenidos por los mismos individuos cuando son examinados con el mismo instrumento en diferentes ocasiones^{54,55}. Un instrumento es fiable cuando los resultados son comparables en situaciones similares⁵⁶. Se puede estimar a través de:

- a) La consistencia interna.- Cuando el instrumento está compuesto por subescalas hay que calcular la correlación entre los ítems que forman cada dominio con la escala total – medida de homogeneidad -. Hay varias técnicas para medirla, pero la más utilizada es mediante el coeficiente alpha (α) de Cronbach, cuyo rango oscila entre el 0 y el 1⁵⁷. Se considera aceptable con valores mayores a 0,7-0,8^{58,59}.
- b) La reproducibilidad test-retest o fiabilidad intra-evaluadores. Cuando es aplicado por los mismos evaluadores, a la misma población y en dos momentos diferentes se obtienen resultados similares – estabilidad temporal –^{60,61}. Para evaluar esta propiedad se puede utilizar el coeficiente de Pearson, Spearman o el índice de

correlación intraclase (ICC), siendo este último el más utilizado⁶⁰. Valores mayores a 0,8 son los adecuados.

A la hora de medir la fiabilidad intraevaluadores también es importante considerar el tiempo transcurrido entre cada medición, que es variable y depende, principalmente, del fenómeno de interés. Es necesario establecer un equilibrio entre ambas mediciones evitando el “efecto aprendizaje”, es decir, que el evaluador pueda recordar las respuestas de la medición anterior. También hay que evitar que las mediciones se distancien mucho en el tiempo para evitar una variación en el fenómeno medido⁶².

- c) La confiabilidad (*agreement*). Se valora mediante la respuesta media estandarizada (RME), que refleja la variabilidad de las puntuaciones de cambio. Valores de 0,20, 0,50 y 0,80 indican instrumentos con escasa, moderada y elevada sensibilidad al cambio respectivamente. Y el cambio mínimo detectable (CMD90).

La validez valora en qué grado un instrumento mide lo que debería medir, es decir, aquello para lo que ha sido diseñado. Se conocen tipos diferentes de validez que pueden medirse en un instrumento. Dependiendo de la escala objeto de validación podrán medirse todas o alguna de estas dimensiones⁶².

- a) Validez de apariencia, aparente o lógica. En este caso no se trata de un elemento estadístico si no más bien un concepto cualitativo que indica que el instrumento de medida parece medir lo que realmente mide⁶³. Normalmente, este es un paso que se lleva a cabo durante el desarrollo de la escala inicial, siendo solo necesario su análisis en aquellos casos donde, durante el proceso de traducción, adaptación cultural y validación, se detecten desajustes⁶⁴.

- b) Validez de contenido: es una medida cualitativa que nos indica si los diferentes ítems incluidos en el instrumento representan adecuadamente los dominios o factores del concepto que se pretende medir^{61,63,65}.
- c) Validez de criterio: se define como el grado de correlación que existe entre un instrumento y otra medida de la variable a estudio que sirva de criterio o referencia^{60,61,66}. Siempre se debe comparar la escala a estudiar con otro instrumento validado que haya demostrado ser válido y fiable para medir lo mismo que mide el instrumento que queremos validar. Sin embargo, cuando no existe un instrumento previamente validado, se puede buscar otra herramienta que sirva de medida comparable. Se establece por medio de coeficientes de correlación de Pearson o de Spearman, dependiendo de las características de distribución de los datos⁶⁷. Para considerar que se ha obtenido un resultado satisfactorio la correlación debe alcanzar un valor $> 0,8$.
- d) Validez de constructo: este tipo de validez sirve de garantía de que las puntuaciones de las respuestas obtenidas pueden ser consideradas y utilizadas como una medición válida del fenómeno a estudiar^{68,69}. Estadísticamente este tipo de validez se mide por medio de un análisis factorial. Este análisis factorial confirmatorio es la herramienta más válida para evaluar la configuración teórica (constructo) subyacente de un instrumento, incluidos sus ítems y factores dentro de una posible estructura jerárquica⁷⁰.
- e) Validez de conclusión estadística: permite medir el grado de confianza que podemos tener. Una forma de medirlo es con los efectos suelo y techo. Hacen referencia al porcentaje de sujetos que obtienen las puntuaciones más bajas y más altas posibles. No debe superar el 15% de los que contestan ya que, en caso contrario, se limita la validez de contenido de la escala porque los pacientes con

valores extremos no pueden distinguirse entre sí. Su presencia hace que los valores extremos puedan perderse indicando una validez de contenido limitada.

Esta propiedad es la que permite realizar las interpretaciones correctas de los resultados obtenidos al aplicar una escala y establecer la relación con la variable que se está midiendo.

La sensibilidad al cambio es la capacidad del instrumento para detectar cambios clínicamente importantes a lo largo del tiempo. El cambio mínimo detectable se mide principalmente a través del tamaño del efecto (TE) o relación entre la media de las diferencias entre las puntuaciones basal y al seguimiento. Y el error mínimo estandarizado (EMM), que evalúa la cantidad de error que uno puede considerar como error de medición. A mayor puntuación menor precisión del instrumento.

El español, después del chino mandarín, es el segundo idioma mas hablado en el mundo y está experimentando un aumento en el número de hablantes. En 2015, casi 470 millones de personas lo usan como lengua nativa (6,7% de la población mundial)⁷¹. No sólo se habla en España e Hispanoamérica, con pequeñas variaciones, sino que también se utiliza de modo habitual en muchos otros países. Por ejemplo, en Estados Unidos de Norteamérica, lo hablan, como lengua materna, más de treinta y seis millones de personas.

Por esto, en España y en países iberoamericanos, ha aumentado la adaptación de instrumentos generados en países anglosajones en lugar de crear instrumentos propios. Además, la dificultad que supone desarrollar nuevos instrumentos de evaluación y la necesidad de comparar los resultados de distintos estudios, hacen necesario traducir y

adaptar los cuestionarios previamente desarrollados y validados en otros países o culturas⁷².

Recientemente se ha publicado una revisión sistemática sobre cuestionarios de hombro autocumplimentados adaptados al portugués⁷³. En 2013⁷⁴ se hizo una revisión de los cuestionarios adaptados al español para pacientes con dolor cervical y lumbar.

No se han publicado revisiones similares de cuestionarios para patologías de las articulaciones más relevantes, como son el hombro, la rodilla y la cadera en español.

Nuestro objetivo ha sido realizar una revisión sistemática, efectuando una búsqueda bibliográfica exhaustiva, sobre los cuestionarios autocumplimentados de hombro, rodilla y cadera adaptados al español, analizando tanto la calidad metodológica del proceso de adaptación transcultural como las propiedades psicométricas de la versión española.

Tras esa primera fase, se creó un grupo de expertos, mediante la técnica de grupo nominal, para llegar a un consenso y seleccionar la escala de hombro más idónea para la evaluación clínica del paciente con patología de esa articulación e incluirla en el sistema de historia clínica electrónica estándar de empleo en los hospitales de la Comunidad Autónoma de Madrid (-*Selene*-) para facilitar su uso en la práctica clínica.

Consenso es, según la RAE, el “acuerdo producido por consentimiento entre todos los miembros de un grupo o entre varios grupos”.

Desde hace años se emplean técnicas de trabajo en grupo en el mundo sanitario^{75,76}, con el fin de disminuir en lo posible los sesgos o influencias en las conclusiones que se extraigan.

Estas técnicas comenzaron a utilizarse como herramientas de ayuda en la planificación y gestión de instituciones o empresas. Pronto pasaron al ámbito de la salud donde se han usado para la realización de predicciones con fines prospectivos, para la detección de necesidades de salud en la comunidad y para el establecimiento de prioridades de actuación o de investigación⁷⁷.

Se recurre al criterio experto cuando no hay suficiente información sobre como orientar las actuaciones clínicas para alcanzar la mayor efectividad clínica, o cuando, por el contrario, la información resulta excesiva y no necesariamente de la mejor calidad.

Los métodos de consenso permiten desarrollar recomendaciones tomando en cuenta la opinión de los expertos y la experiencia de los clínicos, puesto que el resultado del comportamiento conjunto de un grupo de individuos es superior, bajo determinadas condiciones, a la suma de los resultados individuales de los miembros que lo componen^{78,79}.

Si buscamos el descriptor consenso [*consensus*] en PubMed, el repositorio bibliográfico mundialmente más utilizado en biomedicina, se describe como “acuerdo general u opinión colectiva; juicio que alcanza a la mayoría de los interesados”.

En el ámbito de la salud, se han empleado tres abordajes principales de consenso experto⁸⁰: el método Delphi, introducido en los años sesenta^{81,82}, seguido en los setenta por la técnica de Grupo Nominal⁸³ y, a partir de 1977, el método de la Conferencia de

Consenso, desarrollado por el *NIH Consensus Development Program*, del *National Institute of Health* de los EE.UU.⁸⁴.

Estos tres métodos de consenso (Método Delphi, Grupo Nominal y Conferencia de Consenso) utilizan técnicas de reflexión individual y estimulación de la creatividad, algún tipo de discusión estructurada y la combinación de juicios individuales para construir el criterio global de grupo⁸⁵.

Las técnicas de consenso deben considerarse una vía particular con fines propios y estrategias específicas (alcanzar el acuerdo de un conjunto, que no grupo, de personas expertas en un tema científico-técnico, para acordar cuáles son las mejores ideas o las mejores decisiones a tomar sobre un problema planteado).

La primera que se utilizó fue la técnica Delphi, creada en 1952 en los Estados Unidos, por la *Rand Corporation*⁸¹. En su forma original, es una técnica peculiar ya que los participantes del grupo nunca se llegan a reunir. Es más, ninguno conoce quién más participa en él. Con él, se busca acercarse al consenso de un grupo de expertos mediante el análisis y la reflexión de un problema definido.

Se realiza enviando los cuestionarios por correo a un grupo de expertos, para que califiquen una serie de recomendaciones. Estas calificaciones se resumen y reenvían a los expertos, de manera iterativa, para evaluar la posibilidad de que haya cambios en la calificación que permitan llegar a un mayor grado de acuerdo, hasta llegar a una situación de consenso estable para elaborar un informe final de conclusiones. Existen modificaciones a este método (Delphi cara a cara, Delphi telefónico). El número de expertos oscila entre 6 y 30, aunque pueden manejarse grupos más grandes⁸⁶⁻⁸⁹.

Como ventajas destaca la comodidad de poder hacerse sin la presencia física de los expertos y la posibilidad de poder acceder a un amplio número de ellos. A cambio, la elaboración y envío de cuestionarios puede resultar una tarea engorrosa y durar unos 45 días para completar el intercambio de documentos.

El Grupo nominal (o panel de expertos) fue dada a conocer en 1968 como variante de la técnica Delphi⁸³. Su nombre procede de que los miembros del grupo sí conocen los nombres de quienes están participando, (conocen su nombre), al contrario que en el Delphi, en el que permanecen anónimos. La dinámica es sencilla: bajo la coordinación de un facilitador, un grupo de 9 a 12 expertos se reúnen para analizar y calificar una serie de preguntas. Esta calificación se realiza de manera iterativa en dos rondas, buscando reducir la variabilidad de las opiniones, con el fin de alcanzar un acuerdo. Este tipo de técnica se ha utilizado ampliamente para generar recomendaciones en salud⁹⁰.

El Grupo Nominal, adecuadamente conducido, gracias a la presencia de expertos y a las posibilidades de interacción entre ellos, permite profundizar más en las cuestiones que el Delphi y es relativamente más rápido que este, pero anula el efecto de influencias de personajes con estilos dominantes sobre la opinión de otros⁹¹.

Las Conferencias de consenso se diseñaron en respuesta a la complejidad creciente de las cuestiones a resolver planteadas por el mundo sanitario, en el *National Institute of Health*, de Estados Unidos, en 1977⁸⁴. Desde entonces y hasta hoy, se han analizado allí más de 150 temas con esta metodología. Es una técnica que permite una participación amplia de interesados^{92,93}.

El objetivo de las Conferencias de consenso es elaborar recomendaciones para la práctica clínica en una sesión pública donde los expertos presentan la mejor evidencia disponible para responder a las preguntas clínicas planteadas, con base en un análisis crítico previo de la bibliografía, ante un jurado y un público.

Durante la sesión los expertos defienden las conclusiones de las revisiones sistemáticas previamente encargadas a un grupo técnico de revisión bibliográfica e interactúan con el público invitado a la conferencia. Todos ellos pueden expresar opiniones en presencia de un jurado multidisciplinar y multiprofesional que, tras escrutar los datos, escuchar opiniones y clarificar los aspectos necesarios, elabora un informe de conclusiones y recomendaciones en una sesión de puesta en común a puerta cerrada. Para ello, se requiere el acuerdo consensuado de al menos el 80% de los 8-16 miembros que componen dicho jurado, esto es, el acuerdo de una mayoría absoluta cualificada (reforzada) del grupo⁹⁴.

La conferencia de Consenso tiene la ventaja añadida de la participación de los expertos en el propio proceso de elaboración, lo que puede resultar determinante para que sus conclusiones sean aceptadas y seguidas por los clínicos. No obstante, esta metodología puede resultar sesgada por las opiniones preponderantes de los líderes participantes y expuesta a influencias de los financiadores.

La toma de decisiones por consenso depende de que se cumplan ciertas condiciones previas: a) Objetivo común -todo el mundo presente en la reunión tiene que tener un interés común-; b) Compromiso con el consenso -requiere compromiso, paciencia tolerancia y la voluntad de poner al grupo por encima-; c) Tiempo suficiente -pues todas las técnicas de toma de decisiones necesitan tiempo si se quiere que la calidad de la decisión sea alta-; d) Proceso claro -es esencial que todo el grupo comprenda el proceso

que usa la reunión- y e) Buena dinamización y participación activa -la dinamización debe asegurarse de que el grupo funciona armónica, creativa y democráticamente-⁹⁵⁻⁹⁷.

A modo de resumen, la tabla 2 recoge las ventajas e inconvenientes de cada uno de los métodos, valorados críticamente por un experto conocedor de estas metodologías⁹⁸. Como elemento de comparación basal se proponen las características esperables de una reunión informal de expertos donde estos discutiesen de forma no estructurada a la búsqueda de un consenso grupal.

Tabla 2. Comparativa de los principales métodos de consenso. Ventajas e inconvenientes. Reproducida de Peiró S, Protella E. 2003⁸⁸.

Técnicas de consenso	Reuniones no estructuradas	Grupo Nominal	Método Delphi	Conferencia de Consenso
PARTICIPANTES				
Minimización de influencias entre participantes	No	Alta	Muy alta	No
Da iguales oportunidades de participación	No	Sí	Sí	No
Buen funcionamiento con grupos heterogéneos	No	Sí	Sí	Sí
Garantiza la representatividad de participante	No	No	No	Algo
Necesidad de habilidades de comunicación verbal	Sí	Mínima	No	Sí
Necesidad de habilidades de comunicación escrita	No	No	Elevada	No
Sensación de manipulación	Escasa	Alta	Alta	Media
RESULTADOS				
Capacidad de obtención de consenso	Escasa	Alta	Alta	Alta
Capacidad de generación de nuevas ideas	Escasa	Muy alta	Alta	Media
Aportación de perspectivas distintas	Escasa	Muy alta	Alta	Muy alta
Integración de información diversa	A veces	Muy alta	Muy alta	Muy alta
Exploración de más de un problema a la vez	Sí	No	Sí	No
Dificultad en la interpretación de resultados	Alta	Media	Media	Baja
OPERATIVA				
Pérdida de opiniones externas	No	Siempre	Siempre	A veces
Fácil realización	Sí	Sí	No	No
Esfuerzo en la preparación	Bajo	Medio	Alto	Alto
Duración	Corta	Corta	Larga	Media
Necesidad de reunir miembros	Sí	Sí	No	Sí

5. JUSTIFICACIÓN DEL ESTUDIO

En los sistemas de evaluación de los resultados de los tratamientos un aspecto esencial es que el propio paciente, mediante cuestionarios, valore, desde su perspectiva, los resultados alcanzados.

Aunque existen numerosos cuestionarios para medir la limitación funcional percibida por el paciente, la mayoría han sido elaborados en países anglosajones.

Para poder utilizarlos correctamente en otros idiomas, países y/o culturas es necesario, primero, realizar una adaptación transcultural y, después, comprobar que la escala mantiene la fiabilidad y la validez de la versión original.

El español, después del chino mandarín, es el segundo idioma mas hablado en el mundo⁷².

En 2013 se hizo una revisión de los cuestionarios adaptados al español para pacientes con dolor cervical y lumbar⁷⁵.

No se han encontrado revisiones similares de cuestionarios para patologías de las articulaciones más relevantes (hombro, rodilla y cadera) en español.

Por lo expuesto en el párrafo anterior, se propone una revisión sistemática de los cuestionarios autocumplimentados de hombro, rodilla y cadera adaptados al español analizando tanto la calidad metodológica del proceso de adaptación transcultural como las propiedades psicométricas de la nueva versión obtenida, para así saber qué cuestionario es más indicado en una situación concreta considerando criterios metodológicos (propiedades psicométricas) y prácticos (tiempo necesario para cumplimentarla y utilidad para determinadas patologías).

6. HIPÓTESIS

1.- El procedimiento de adaptación transcultural y de las propiedades clinimétricas de los cuestionarios autocumplimentados de valoración funcional para afecciones de las articulaciones más relevantes (hombro, rodilla y cadera) adaptados al español ha seguido los criterios de calidad aceptados.

2.- Es posible elegir una escala de valoración de hombro que presente las mejores propiedades instrumentales y capacidad predictiva y de evaluación para su utilización en la práctica clínica entre las escalas validadas al español.

7. OBJETIVOS

1. Revisar sistemáticamente los cuestionarios autocumplimentados de hombro adaptados al español analizando tanto la calidad metodológica del proceso transcultural como las propiedades psicométricas de la nueva versión obtenida.
2. Revisar sistemáticamente los cuestionarios autocumplimentados de rodilla adaptados al español analizando tanto la calidad metodológica del proceso transcultural como las propiedades psicométricas de la nueva versión obtenida.
3. Revisar sistemáticamente los cuestionarios autocumplimentados de cadera adaptados al español analizando tanto la calidad metodológica del proceso transcultural como las propiedades psicométricas de la nueva versión obtenida.
4. Elección del cuestionario de hombro más indicado para su uso en la práctica clínica.

8. MATERIAL Y MÉTODOS

Para conseguir los objetivos planteados en esta investigación, se ha procedido a trabajar en dos fases diferenciadas.

En la primera fase se ha realizado una revisión sistemática de la adaptación transcultural al español de los cuestionarios autocomplimentados de valoración del paciente con las afecciones más frecuentes del aparato locomotor, que afecten a hombro, rodilla y cadera.

En la segunda fase se procedió a localizar y conocer las escalas adaptadas para la articulación del hombro. Mediante consenso de expertos, a través de grupos nominales, se ha seleccionado el cuestionario de hombro considerado como el más idóneo para la evaluación del paciente en la práctica clínica, siendo incorporado al sistema informático del hospital, facilitando así su uso.

Cabe señalar que esta segunda fase el estudio se ha centrado en la articulación del hombro porque los motivos por los que los pacientes acuden a consulta más frecuentemente se localizan en esta zona anatómica^{5,7}.

A continuación, se describirán detalladamente las dos fases mencionadas

8.1.- Primera fase

8.1.1. Estrategia de búsqueda

Se ha realizado una revisión sistemática de los artículos publicados sobre la adaptación transcultural al español de cuestionarios autocomplimentados para valorar a

pacientes con afecciones de hombro, rodilla y cadera y sobre su proceso de validación a nuestro idioma.

Se efectuó una búsqueda bibliográfica en las principales bases de datos biomédicas internacionales desde la fecha que figura como su creación (MEDLINE, 1966; EMBASE, 1974; CINAHL, 1997 y *Web of Science*, 1900) hasta el 31 de diciembre de 2015 para hombro, 31 de mayo de 2016 para rodilla y julio de 2016 para cadera. Además, se revisaron otras bases de datos y directorios con predominio de bibliografía en español o exclusivamente en este idioma: BIREME (Scielo, Lilacs e IBECs), MEDES, AMED, IME y DIALNET.

Los términos utilizados y la estrategia de búsqueda en MEDLINE fueron los siguientes: “Outcome” or “Questionnaire” or “Score” or “Functional Scale” or “Assessment tool” or “Instruments” and “Shoulder” or “Rotator cuff” or “Disability” or “Upper limb” or “Arm” (para hombro); «knee» or «disability» or «lower limb» (para rodilla); «Hip» or «Lower limb» (para cadera) and “Spanish” or “Spanish version” or “Spanish cultural adaptation” or “Spanish Assessment” or “Spanish traslation” or “Cross-cultural adaptation” or “Validation”.

Los términos empleados para buscar en las otras bases de datos fueron similares.

Se realizó, también, una búsqueda manual utilizando los nombres de las diferentes escalas de hombro como palabra clave en las distintas bases de datos.

Además, se efectuó una búsqueda en la web, incluyendo *Google Scholar*, con el fin de poder abarcar otro tipo de publicaciones, incluyendo literatura gris.

Finalmente se examinaron manualmente las referencias de los artículos obtenidos.

8.1.2. Criterios de selección

Dos autores revisaron el título y el resumen de cada artículo obtenido.

Si su lectura sugería que podía ser elegible se realizó la lectura completa del artículo.

8.1.2.1.- Criterios de inclusión

Se incluyeron artículos, sin restricción idiomática, sobre estudios observacionales que describieran el proceso de adaptación transcultural al español de cuestionarios autoaplicados para realizar la valoración funcional de pacientes con afecciones de hombro, rodilla y cadera, y que analizaran, también, las propiedades psicométricas de la nueva versión.

8.1.2.2.- Criterios de exclusión

Se excluyeron los artículos donde sólo se analizaran las propiedades de un cuestionario previamente adaptado, protocolos de investigación y resúmenes de congresos.

En las figuras 2, 3 y 4 se muestran los diagramas de flujo del proceso de selección de los artículos de cada articulación: hombro, rodilla y cadera respectivamente.

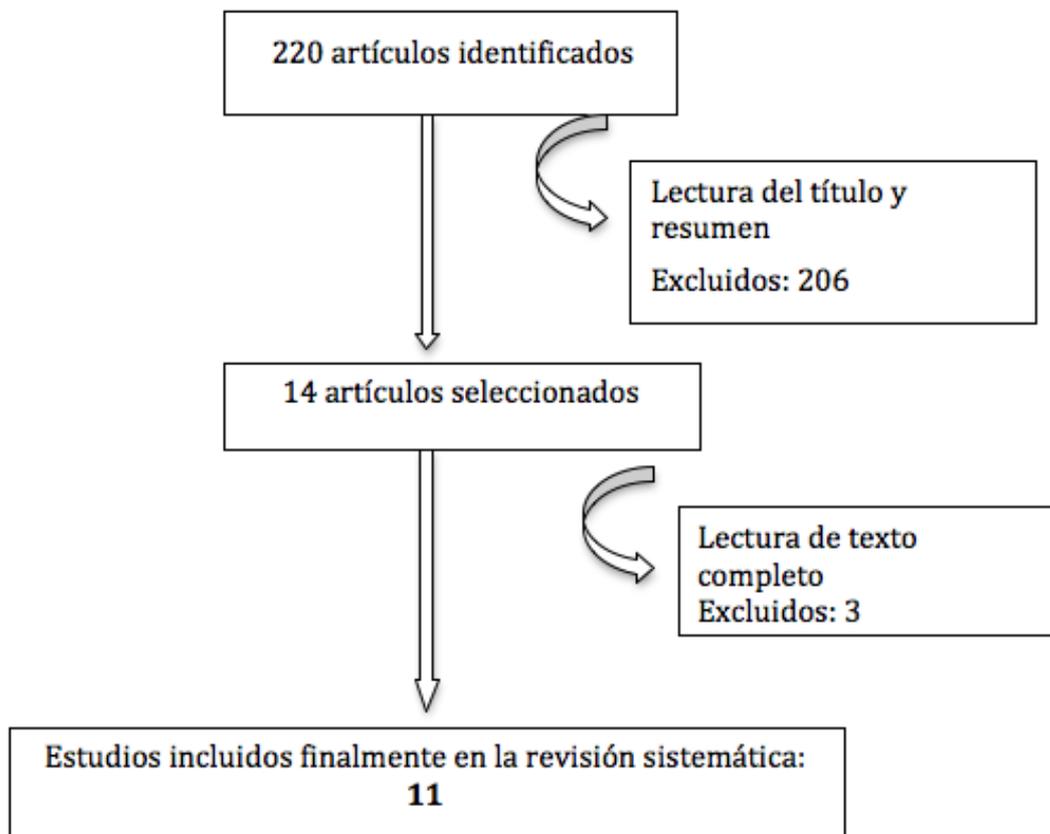


Figura 2. Diagrama de flujo que muestra la selección de los estudios incluidos para patologías de hombro.

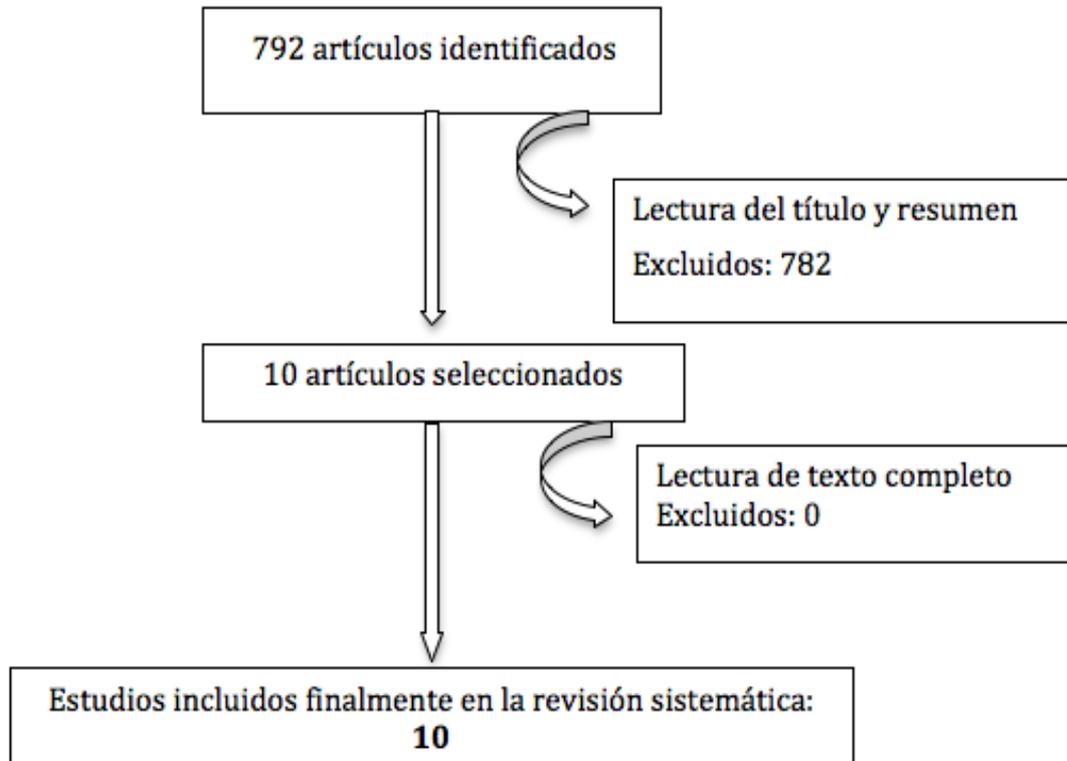


Figura 3. Diagrama de flujo que muestra la selección de los estudios incluidos para patologías de rodilla.

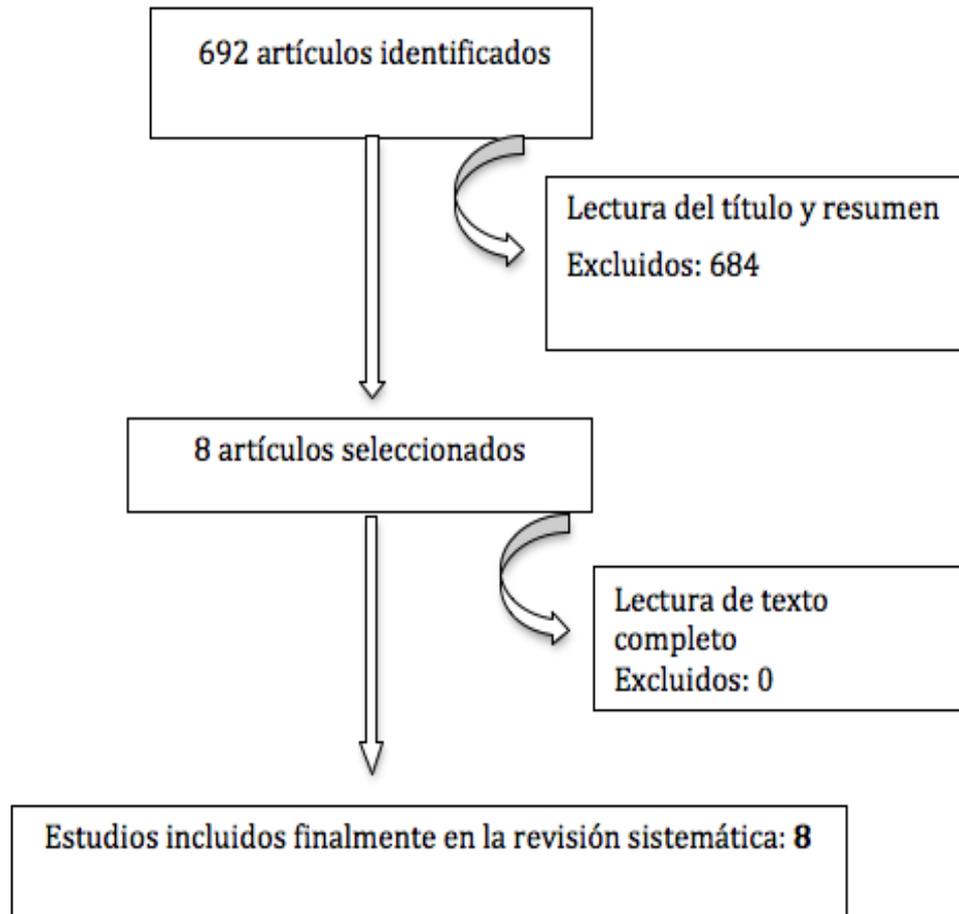


Figura 4. Diagrama de flujo que muestra la selección de los estudios incluidos para patologías de cadera.

8.1.3. Análisis de la información

A partir de los artículos seleccionados, dos autores estructuraron y analizaron los resultados de cada uno de los manuscritos.

En caso de discrepancia se buscó el consenso y, cuando fue necesario, se recurrió a un tercer revisor.

El análisis de la información incluyó:

8.1.3.1. Características de los participantes en el estudio.

Se recogieron los datos de los pacientes que formaban parte del estudio de cada uno de los cuestionarios: país y localidad, número total de pacientes incluidos, diagnóstico, edad, sexo y tiempo que tardaron en cumplimentar el cuestionario.

Se comprobó que incluyeran al menos 50 pacientes: número mínimo recomendado para los estudios de adaptación transcultural^{100,101}.

8.1.3.2. Evaluación de la metodología utilizada para efectuar la adaptación transcultural.

Se comprobó si se realizaron los cinco pasos habitualmente recomendados en la bibliografía internacional^{49,53,65,102-104} y si se siguieron correctamente.

Los cinco pasos son:

- 1) Traducción directa del cuestionario original al español.

- 2) Síntesis de traducciones y solución de posibles discrepancias entre los traductores con algún miembro del equipo de investigación.
- 3) Traducción inversa o retro-traducción.
- 4) Revisión por un comité de expertos.
- 5) Prueba piloto (pre-test) del cuestionario pre-final con sujetos de habla española.

Se consideró si cada uno de los pasos se había hecho correctamente (o si era dudosa la corrección), si no se había realizado o si los autores no facilitaban la información necesaria.

La figura 5 muestra la síntesis de la metodología para la traducción y adaptación cultural de las escalas adaptadas de Beaton et al⁵³.

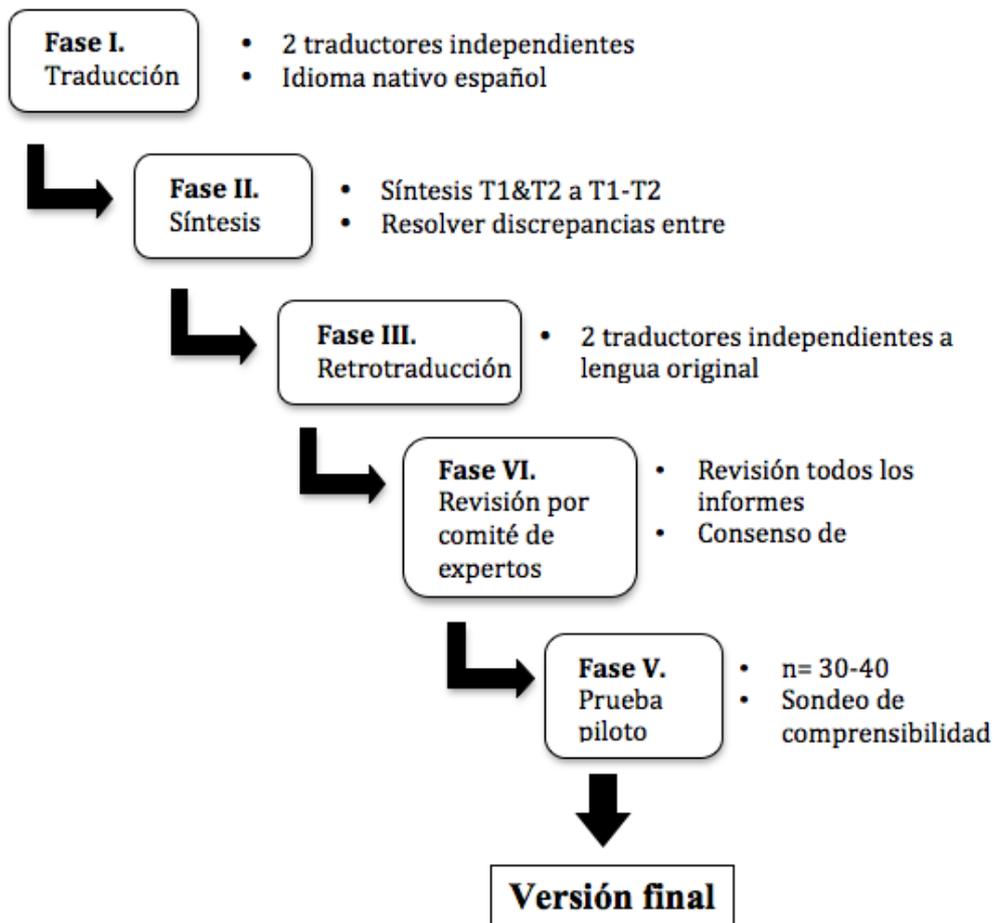


Figura 5. Síntesis de la metodología para la traducción y adaptación cultural de las escalas adaptado de Beaton et al.

8.1.3.3. Evaluación de las propiedades psicométricas de la versión española.

Se analizaron los siguientes aspectos: fiabilidad, validez y sensibilidad al cambio^{49,101,104-107}.

La fiabilidad, propiedad que evalúa la constancia o precisión de los resultados que obtiene el instrumento, se midió mediante:

- La consistencia interna o homogeneidad del instrumento a través del alfa de Cronbach.
- La fiabilidad intraevaluador o test-retest, mediante el ICC y el coeficiente de Spearman.
- La confiabilidad, mediante el RME, que refleja la variabilidad de las puntuaciones de cambio, y el CMD90.

La validez, propiedad que valora en qué grado un instrumento mide lo que debería medir, es decir aquello para lo que fue diseñado, se estudió a través de:

- Validez de criterio o grado de correlación que existe entre un instrumento y otra medida de la variable a estudio que sirva de criterio o referencia, mediante los coeficientes de correlación de Pearson o de Spearman, considerando satisfactorio un resultado $> 0,8$.
- Validez de conclusión estadística o grado de confianza que podemos tener, medido con los efectos suelo y techo. Su presencia hace que los valores extremos puedan perderse indicando una validez de contenido limitada.

La sensibilidad al cambio, capacidad del instrumento para detectar cambios clínicamente importantes a lo largo del tiempo, se ha medido a través del TE para calcular el CMD y el EMM, para saber la cantidad de error que uno puede considerar como error de medición.

8.1.3.4. Evaluación de la aplicabilidad directa de las escalas adaptadas a un país hispanohablante para ser aplicada en otro.

Se analizó si las adaptaciones realizadas en países latinoamericanos necesitaban alguna modificación para poder utilizarse en España y viceversa.

Cuando fue necesario se estableció contacto con los autores, mediante correo electrónico, para aclarar dudas.

8.2. Segunda fase

En la segunda fase se procedió a localizar y conocer las escalas adaptadas para la articulación del hombro. Mediante consenso de expertos, a través de grupos nominales, se seleccionó el cuestionario de hombro considerado más idóneo para la evaluación del paciente en la práctica clínica, incorporándose al sistema informático del hospital para facilitar su uso.

8.2.1. Selección de expertos

Partiendo de las competencias seleccionadas para el estudio y considerando las funciones definidas para los profesionales que se han propuesto se plantea un panel de expertos.

Para la selección de los expertos que participaron en esta fase, se toma como referencia la premisa de la técnica de grupo nominal⁸⁴, según la cual son los profesionales expertos los que mejor pueden describir y definir su trabajo basándose en la forma de realizar sus tareas y teniendo en cuenta que dichas tareas requieren la utilización adecuada de conocimientos, habilidades y destrezas además de actitudes positivas para facilitar el desempeño con éxito.

En cuanto a los procedimientos de elección de los expertos, los autores¹⁰⁸⁻¹¹⁰ indican una diversidad que incluye desde los que no implican ningún filtro de selección, como en los casos de afinidad o cercanía entre el experto y el investigador, hasta los que utilizan una serie de criterios estructurados como son el biograma o el coeficiente de competencia experta.

El primero de los dos criterios en el que nos hemos basado para nuestro trabajo es la afinidad y cercanía entre los expertos e investigadores, seleccionando a profesionales médicos de las distintas especialidades que con más frecuencia valoran las patologías de hombro.

Para cumplir con el segundo, el coeficiente de competencia experta, se parte de las personas que inicialmente se han considerado expertos para que con su opinión y autovaloración indiquen su nivel sobre el conocimiento acerca del objeto de investigación, así como de las fuentes que les permiten argumentar y justificar dicho nivel¹⁰⁸.

Además de los criterios a utilizar para su selección, otro de los problemas con el que nos encontramos es el del número de expertos necesarios para llevar a cabo un estudio. Aquí hallamos diferentes propuestas: Malla y Zabala¹⁰⁹ sugieren que su número debe oscilar entre 15 y 20; Landeta¹¹⁰ indica que debe estar comprendido entre 7 y 30; Cabero y Llorente¹⁰⁸ sugieren que el número de expertos depende de aspectos como la facilidad para acceder a ellos o la posibilidad de conocer expertos suficientes sobre la temática objeto de la investigación.

Atendiendo a consideraciones éticas, cabe señalar que todos los participantes fueron debidamente informados sobre el objetivo y la utilización de la información que se pretendía obtener.

8.2.2. Criterios de selección de los expertos

- Estar en posesión del título de licenciado en Medicina.
- Estar en activo en cualquier ámbito de la Medicina.
- Tener al menos cinco años de experiencia como médico adjunto en el campo de Cirugía Ortopédica y Traumatología, Reumatología o Medicina Física y Rehabilitación.

- Trabajar en el Hospital Universitario Fundación de Alcorcón.
- Tener experiencia en la patología del hombro.

8.2.3. Convocatoria

Se les convocó vía email para una reunión. En él, se adjunta una carta de invitación a dicho encuentro (*Anexo I*) donde además se les explica el objetivo a conseguir (seleccionar la escala de hombro más apropiada para utilizar en la práctica clínica e incorporarla en el sistema informático del hospital –“Selene”-).

Se adjuntó a la convocatoria, como documentación, la información sobre las escalas adaptadas al español para la valoración funcional de hombro adaptadas a la población española sobre las que se va a trabajar¹¹¹.

Además, se entregaron las escalas como apoyo al trabajo para facilitar así el análisis de la calidad del proceso de adaptación y las propiedades de cada versión.

Se les solicitó que revisasen esta información para el día de la reunión y dieran su consentimiento para participar.

8.2.4. Reunión de Expertos

- Fecha: jueves 17 de marzo de 2016, las 14:00.

- Lugar: Sala de reuniones del departamento de Medicina Física y Rehabilitación del Hospital Universitario Fundación Alcorcón.
- Duración: 1 hora y 30 minutos.

8.2.5. Participantes

Una vez identificados los criterios de selección que hemos señalado anteriormente, nos pusimos en contacto con diferentes personas que creíamos que cumplían los criterios anteriores y que conocíamos por participación en eventos, colaboración en proyectos anteriores de investigación o con los que se tenía alguna vinculación histórica.

De todas las cartas de enviadas a los expertos solicitando su participación en la investigación, contestaron 20 personas, que fueron invitadas a la reunión. De todas ellas, finalmente, acudieron 12 expertos (Tabla 3).

Como la clave para tomar decisiones por consenso en un grupo grande es reducir el problema al caso en el que el consenso sí que es posible, dividimos el grupo en dos grupos pequeños para facilitar el diálogo, el debate y presentar sus conclusiones de forma clara y ordenada al otro grupo. Este procedimiento ha de ser lo más claro posible para todos y debe aceptarse por todos los participantes.

Junto a las reflexiones metodológicas no hay que perder de vista que sobre todo es la práctica y la experiencia de los participantes lo que hace posible llegar a una decisión por consenso.

Se decidió dividir al grupo de 12 expertos en dos equipos, uno formado por seis médicos de Medicina Física y Rehabilitación y el otro por el resto de especialidades (cuatro de Cirugía Ortopédica y Traumatología y dos de Reumatología) con el objetivo de obtener una visión diferente a los especialistas de rehabilitación.

Tabla 3.- Participantes en el panel de expertos.

	Convocados	Asistentes
Traumatólogos	10	4
Reumatólogos	4	2
Rehabilitadores	6	6
Total	20	12

8.2.6. Desarrollo de la reunión

El desarrollo de la jornada, se dividió en dos fases, tal como se describe a continuación (figura 6):

- Fase 1:

1.1. Introducción, a modo de resumen, de la información enviada mediante una exposición de 5 minutos.

El moderador presenta el planteamiento del estudio y explica los conceptos generales de las distintas escalas y sus elementos.

Se informa sobre los objetivos: la evaluación de las escalas de hombro y la elección final de una de las escalas para introducirla en el sistema informático del hospital (“Selene”).

1.2. Explicación de la dinámica del grupo por parte del moderador (5 minutos).

El moderador explica en qué consiste un grupo nominal y las principales propiedades a estudiar de cada escala. Se divide a los participantes en dos grupos de 6 expertos cada uno.

A continuación, se proponen los siguientes ejercicios:

- Ejercicio 1: Se debe reflexionar sobre las diferentes escalas de hombro adaptadas al español, estudiando las características principales de cada una de ellas.

Principales características a estudiar:

- Correcta adaptación a la población española.
 - Viabilidad (sencillez, aceptable para el paciente y el profesional).
 - Fiabilidad (precisión).
 - Validez (mide lo que se propone medir);
 - Reproducibilidad (intra- e inter- observador);
 - Sensibilidad a los cambios clínico (capacidad para detectar cambios importantes).
- Ejercicio 2: Elegir un total de 4 escalas de hombro para estudiarlas más a fondo con el fin de elegir en la siguiente fase la escala definitiva.

1.3. Desarrollo del trabajo de grupos (30 minutos)

Para el desarrollo de los trabajos en grupo se repartieron los distintos objetivos a conseguir.

Se entregó como documentación las once publicaciones relevantes, con las doce adaptaciones de las nueve escalas de valoración funcional de hombro adaptadas a la población española encontradas sobre las que se iba a trabajar.

Además, se entregaron las escalas como apoyo para el trabajo para facilitar así el análisis de la calidad del proceso de adaptación y las propiedades de cada versión.

La dinámica de grupo comenzó por realizar una tabla comparativa con las propiedades psicométricas a estudiar de todas las escalas, para proceder después a la selección de los elementos competenciales a estudiar. Posteriormente, estas propiedades se agruparon, suprimiéndose aquellas cuyos contenidos se superponían o no eran lo suficientemente importantes. Por último, se recogieron las conclusiones del grupo sobre los elementos y componentes trabajados.

1.4. Puesta en común de las conclusiones. Cada uno de los grupos dispuso de 10 minutos para argumentar su exposición. (20 minutos).

- Fase 2:

2.1. Nuevo desarrollo del trabajo. Decisión final. (10 minutos).

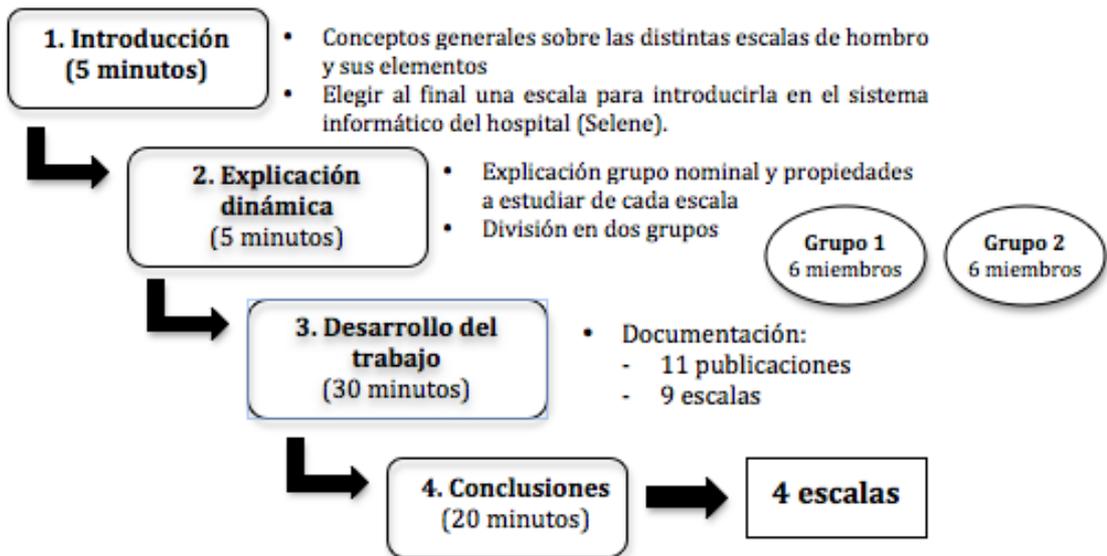
Tras la selección de cuatro escalas se vuelven a dividir en los grupos de trabajo para seleccionar la escala definitiva más completa para valorar las patologías de hombro.

El examinador hizo un resumen a los participantes y solicitó que priorizaran la escala que les parecía más importante.

2.2. Consenso final. (20 minutos).

Se solicitó a los expertos que realizaran una nueva discusión acerca de la selección de la escala elegida para incorporar al sistema informático del hospital, pero sólo para perfilar o aclarar los conceptos. Se trataba de hacer una charla de aclaración, no de lograr un consenso donde pueden jugar su papel las presiones.

• Fase 1:



• Fase 2:

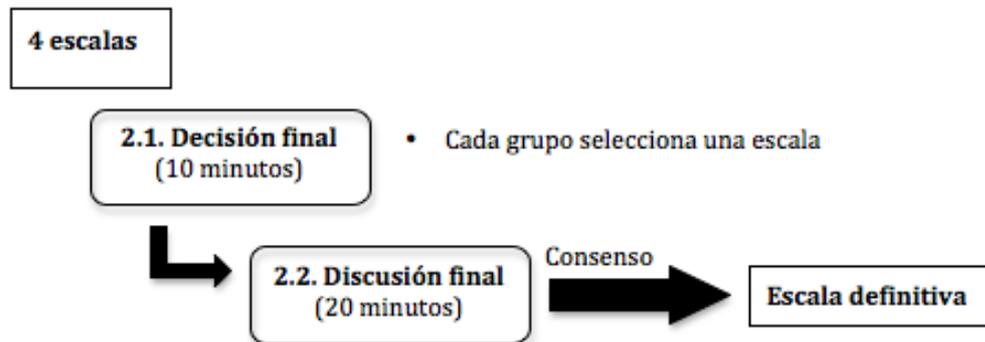


Figura 6.- Desarrollo de la reunión del panel de expertos.

9. RESULTADOS

9.1.- PRIMERA FASE

9.1.1. HOMBRO

Mediante la búsqueda bibliográfica realizada en las principales bases de datos biomédicas identificamos once artículos relevantes¹¹²⁻¹²².

En MEDLINE identificamos ocho publicaciones¹¹²⁻¹¹⁹ y en el resto de las bases de datos encontramos los tres restantes¹²⁰⁻¹²².

En los once artículos se encontraron nueve cuestionarios autocumplimentados traducidos y adaptados transculturalmente a la población española y en los que se habían estudiado las propiedades clinimétricas de la nueva versión. En todos los casos el cuestionario original se había desarrollado en inglés.

Había dos cuestionarios aplicables a cualquier tipo de afección del miembro superior: *Disabilities of the Arm, Shoulder and Hand Questionnaire* (DASH)¹¹²⁻¹¹⁴ y *Upper Limb Functional Index* (ULFI)¹¹⁵; cuatro cuestionarios generales de hombro: *Simple Shoulder Test* (SST)¹¹⁶, *Shoulder Pain and Disability Index* (SPADI)^{117,118}, *Oxford Shoulder Score* (OSS)¹¹⁸ y *Shoulder Disability Questionnaire* (SDQ)¹¹⁹; y tres cuestionarios para situaciones específicas: *Western Ontario Rotator Cuff Index* (WORC)¹²⁰ para sujetos con patología manguito rotador, *Western Ontario Shoulder Instability Index* (WOSI)¹²¹ para pacientes con inestabilidad de hombro y *Wheelchair Users Shoulder Pain Index* (WUSPI)¹²² para usuarios de sillas de ruedas.

El cuestionario DASH se adaptó y validó en tres ocasiones, dos en España^{112,113} y una en Puerto Rico¹¹⁴.

El cuestionario SPADI se adaptó y validó en dos ocasiones, ambas en España^{117,118} y prácticamente de forma simultánea. Una de las dos adaptaciones se realizó conjuntamente con el cuestionario OSS y se publicaron en un mismo artículo¹¹⁸.

En total hubo doce adaptaciones para nueve cuestionarios. Hay que señalar que nuestra búsqueda no encontró ningún cuestionario desarrollado originalmente en español.

En la tabla 4 se pueden ver las características demográficas y clínicas de la población en la que se realizó cada uno de los doce estudios en los once artículos.

La escala adaptada, junto con instrucciones para el paciente, se incluye en la publicación en todos los casos salvo para tres cuestionarios: la adaptación del DASH en Puerto Rico¹¹⁴, OSS¹¹⁸ y WORC¹²⁰.

La mayoría de los artículos se han publicado muy recientemente: seis de las adaptaciones se publicaron en 2015^{116-118,120,121} y una en 2013¹¹⁵. La primera escala adaptada fue el DASH, en 2002, por Rosales et al¹¹² en Tenerife. Hubo cuatro escalas adaptadas en países latinoamericanos^{114,119-121} y el resto se realizó en España^{112,113,115-118,121}. En ocho de las doce adaptaciones^{113,115-118,120,121} se incluyó un número de pacientes superior al recomendado de 50 para este tipo de estudios. Sólo una de las tres adaptaciones de la escala DASH, la realizada por Hervás et al¹¹³ en Valencia en 2006 incluía pacientes con afecciones de hombro. Las otras dos fueron validadas en pacientes con síndrome del túnel carpiano¹¹² y, la versión para Puerto Rico, en pacientes tras cirugía por cáncer de mama¹¹⁴. Los autores que adaptaron simultáneamente los cuestionarios SPADI y OSS incluyeron sólo pacientes intervenidas de cáncer de mama¹¹⁸. En la otra adaptación del SPADI¹¹⁷ se incluyeron pacientes con patología de hombro. En la adaptación de la escala WUSPI se incluyeron sólo pacientes lesionados medulares usuarios de silla de ruedas¹²².

Sólo para cuatro cuestionarios se recogió el tiempo que se tardaba en cumplimentar^{113,120,122} y fue similar al de la versión original.

Tabla 4. Características demográficas y clínicas de la población de los estudios de adaptación transcultural.

ESCALA	Autor (año de la adaptación)	País (Población)	Nº de pacientes incluidos	Diagnóstico	Edad	% de mujeres
DASH ¹¹³	Hervás et al (2006)	España (Valencia)	98	Diversas patologías de miembro superior (73,5 % con problemas de hombro: 34,6% síndrome subacromial, 27,3% tendinitis y 9,1% fracturas humerales o de escápula)	Rango: 27 -76 (mas del 50%: 25-65)	64'3 %
DASH ¹¹⁴	Mulero-Portela et al (2009)	Puerto Rico (San Juan)	44	Tras cirugía de cáncer de mama (operadas en los tres años previos y que habían completado su tratamiento)	Media: 52'59 Rango: 34-84	100%
DASH ¹¹²	Rosales et al (2002)	España (Tenerife)	42	Síndrome del túnel del carpo	Media: 54 Rango: 34-63	85'7%
ULFI ¹¹⁵	Cuesta-Vargas y Gabel (2013)	España (Málaga)	126	Diversas patologías del miembro superior (56 % con problemas de hombro: síndrome subacromial o tendinitis)	Media: 49 (DE ± 21)	54'8 %
SST ¹¹⁶	Membrilla-Mesa et al (2015)	España (Granada)	66	Dolor de hombro por diversas patologías (31 % fractura humeral, 21,23% tendinitis calcificante y 19,74% rotura del manguito rotador)	Media: 51 (DE ± 12)	59'13 %
SPADI ¹¹⁷	Membrilla-Mesa et al (2015)	España (Málaga)	136	Dolor de hombro por diversas patologías (22,1% tendinopatía bicipital, 21,3 % fractura humeral, 17,6% rotura del manguito rotador y 11,8% tendinitis calcificante)	Media: 49'8 (DE ± 15)	55'1%
SPADI ¹¹⁸	Torres-Lacomba et al (2015)	España (Alcalá de Henares)	120	Tras cirugía de cáncer de mama (con menos de seis meses de evolución)	Media: 54'2 (DE ± 11)	100 %
OSS ¹¹⁸	Torres-Lacomba et al (2015)	España (Alcalá de Henares)	120	Tras cirugía cáncer de mama (con menos de seis meses de evolución)	Media: 54'2 (DE ± 11)	100 %
SDQ ¹¹⁹	Álvarez-Nemegyei et al (2005)	Méjico (Yucatán)	35	Dolor de hombro por síndrome subacromial	Media: 55 (DE ± 9)	77'1%
WORC ¹²⁰	Arcuri et al (2015)	Argentina (Buenos Aires)	60	Tendinitis o rotura del manguito rotador	Media: 57 (DE ± 12'3) Rango: 19-76	44 %
WOSI ¹²¹	Arcuri et al (2015)	Argentina (Buenos Aires)	60	Inestabilidad de hombro	Media: 40'12 (DE ± 17) Rango: 17-65	45%
WUSPI ¹²²	Arroyo-Aljaro y González-Viejo (2009)	España (Barcelona)	42	Lesionados medulares (28,9% nivel cervical, 36,8% nivel dorsal alto y 34,2% nivel dorsal bajo)	Mediana: 40 Rango: 15-69	19 %

El cuestionario DASH tiene dos módulos opcionales adicionales, cada uno de cuatro ítems, que fueron adaptados en la versión de Hervás et al¹¹³ y en la realizada en 2009 para pacientes puertorriqueños¹¹⁴. En cada uno de los cuestionarios originales viene especificado claramente el periodo de tiempo, anterior al momento de la cumplimentación, que el paciente tiene que considerar cuando lo contesta. Sin embargo, en las adaptaciones de los cuestionarios WUSPI¹²² y en una de las adaptaciones del SPADI, la de Torres-Lacomba et al¹¹⁸, no se aclara, en la escala adaptada, el lapso de tiempo que debe analizar el paciente al valorar la situación de su hombro. En general el proceso de adaptación transcultural de todas las escalas requirió modificaciones, aunque de poca importancia, respecto a la versión original. En las tres adaptaciones del DASH¹¹²⁻¹¹⁴ no se obtuvieron cuestionarios idénticos. En la versión para Puerto Rico¹¹⁴ fue necesario incluir algunos anglicismos al ser en ese país el inglés la segunda lengua oficial. Tampoco resultaron exactas las dos adaptaciones del SPADI^{117,118} a pesar de que ambas fueron realizadas en España. No obstante, en los dos cuestionarios las diferentes variantes obtenidas fueron equivalentes.

Los principales resultados de la evaluación de la metodología utilizada para la adaptación transcultural se analizan en la tabla 5.

Las tres versiones del cuestionario DASH¹¹²⁻¹¹⁴, las dos adaptaciones del SPADI^{117,118} y las versiones del OSS¹¹⁸, WORC¹²⁰ y WOSI¹²¹ cumplieron rigurosamente los cinco pasos de las recomendaciones de las guías internacionales. Los cuestionarios SDQ¹¹⁹ y WUSPI¹²² fueron los menos estrictos en ese aspecto. La prueba piloto se realizó para ocho escalas^{112-114,117,118,120,121} pero sólo en tres^{112,114,117} se hizo con más de 30 sujetos.

Tabla 5. Evaluación de la metodología empleada para la adaptación transcultural de los cuestionarios.

ESCALA	Traducción	Síntesis	Traducción inversa	Análisis por comité de expertos	Prueba piloto (número de pacientes)
DASH ¹¹³ Hervás et al (2006)	+	+	+	+	+ (15)
DASH ¹¹⁴ Mulero-Portela et al (2009)	+	+	+	+	+ (32)
DASH ¹¹² Rosales et al (2002)	+	+	+	+	+ (50)
ULFI ¹¹⁵	+	+	+	?	-
SST ¹¹⁶	+	+	+	?	-
SPADI ¹¹⁷ Membrilla-Mesa et al (2015)	+	+	+	+	+ (40)
SPADI ¹¹⁸ Torres-Lacomba et al (2015)	+	+	+	+	+ (20)
OSS ¹¹⁸	+	+	+	+	+ (20)
SDQ ¹¹⁹	-	0	-	0	-
WORC ¹²⁰	+	+	+	+	+ (10)
WOSI ¹²¹	+	+	+	+	+ (10)
WUSPI ¹²²	?	0	?	0	-

Las principales propiedades psicométricas analizadas en las versiones españolas se muestran en la tabla 6.

Ningún trabajo evaluó todas las propiedades métricas de la nueva versión. En general la conclusión de los autores de cada una de las adaptaciones realizadas fue que las propiedades psicométricas evaluadas resultaban aceptables y comparables a las de las versiones originales y a otras versiones adaptadas a otros idiomas.

Tabla 6. Principales propiedades psicométricas analizadas en los cuestionarios adaptados.

ESCALA	Consistencia Interna (α de Cronbach) total	Reproductibilidad Test-retest: N° participantes (tiempo entre dos evaluaciones) ICC	Instrumento de comparación: Validez (coeficiente de correlación) (valores negativos indican correlación inversa)	Efectos suelo / techo	Sensibilidad	
					ES	SRM
DASH ¹¹³ Hervás et al (2006)	0.96	30 (de 7 a 10 días) 0.96, $p < 0.01$	- SF-36: subescala función física: -0.509 y subescala dolor: -0.693. Correlación significativa ($p < 0.01$) con todas las dimensiones de SF-36 en el sentido esperado según la hipótesis de partida.	---	1'03	0'86
DASH ¹¹⁴ Mulero-Portela et al (2009)	0'97 (rango entre ítems: 0'44-0'85)	No analizada	- FACT-B: (rango 0'096-0'682). Correlación significativa ($p < 0.01$) con cuatro de las cinco subescalas de FACT-B	No / No	---	---
DASH ¹¹² Rosales et al (2002)	0'95	42 (7 días) 0'97(95% IC: -1'98 a 0'52)	No utilizados (no calculada)	---	---	---
ULFI ¹¹⁵	0.94 (rango entre los ítems: 0'92-0'96)	35 (7 días) 0.93 (rango 0'92-0'95)	- EQ-5D-3L: -0.59 (para la subescala dolor: -0'51) - QuickDASH: No válido (mas del 30% de cuestionarios con excesivas respuestas perdidas) Análisis de estructura factorial: unidimensional	---	---	---
SST ¹¹⁶	0.793 (rango entre los ítems: 0'667-0'723)	21 (2 días) 0.912 (rango 0'687-0'944)	- DASH: -0.73 - EVA dolor: -0.53 - SF-12: subescala función física - 0.47 y subescala mental - 0.43, con $p < 0.001$ Análisis de la estructura factorial: tridimensional	No / No	---	---
SPADI ¹¹⁷ Membrilla-Mesa et al (2015)	0'92 (95% IC: 0'91 a 0'95) y 0'82 (95% IC: 0'76 a 0'86) para cada uno de los dos factores	56 (2 días) 0'91 (95% IC: 0'88 a 0'94) con rango para cada ítem de 0'89 a 0'93	Correlaciones significativas ($p < 0.01$) en el sentido esperado según la hipótesis de partida - DASH: subescala dolor 0'80 y subescala discapacidad 0'76 - SF-12: total 0'40 (con las subescalas de función física y mental) - SST: subescala dolor -0'71 y subescala discapacidad -0'75 - EVA: subescala dolor 0'67 y subescala discapacidad 0'65 Análisis de estructura factorial: bidimensional	---	---	---
SPADI ¹¹⁸ Torres-Lacomba et al (2015)	0.965 (0'931 en subescala dolor y 0'953 en subescala función)	20 (2 días) 0.992 (0'986 en subescala dolor y 0'991 en subescala función) $p < 0.01$	- OSS: -0.674 - FACT-B: - 0.298 - SDQ: 0.432 - SF-36: -0'452 a -0'315 (para dimensiones función física, rol físico, dolor corporal y rol emocional) con $p < 0.01$	No / No	Escala global 0'59	0'82 Subescala dolor 0'59 0'78 Subescala función 0'82 1'13 $p < 0.0001$

Tabla 6 (continuación). Principales propiedades psicométricas analizadas en los cuestionarios adaptados.

ESCALA	Consistencia Interna (α de Cronbach) total	Reproductibilidad Test-retest: N° participantes (tiempo entre dos evaluaciones) ICC	Instrumento de comparación: Validez (coeficiente de correlación) (valores negativos indican correlación inversa)	Efectos suelo / techo	Sensibilidad ES SRM
OSS ¹¹⁸	0.947	20 (2 días) 0.974	- SPADI: -0.674 (subescala dolor -0.640 y subescala discapacidad -0.645) - FACT-B: -0.343 - SDQ: -0.469 - SF-36: 0.312 a 0.391 (para función física, rol físico, dolor corporal y rol emocional) con $p < 0.01$	No / No	-0.5 - 0.70 $p < 0.0001$
SDQ ¹¹⁹	0.99	35 (1 hora) 0.99 (95% IC: 0.998 a 0.999)	No utilizados (no calculada)	---	---
WORC ¹²⁰	0.952 (rango entre los 5 dominios: 0.824-0.953)	10 (de 14 a 21 días, mediana 14) 0.909 (95% IC: 0.841 a 0.949)	- SST: 0.756 - Escala de Constant: 0.60	No / No	---
WOSI ¹²¹	0.948 (rango entre los 4 dominios: 0.783-0.906)	60 (de 14 a 21 días, mediana 14) 0.851 (95% IC: 0.587 a 0.879)	- SST: 0.498 - Escala de Constant: 0.60	No / No	---
WUSPI ¹²²	0.88	8 (No especificado) 0.96	- EVA dolor (en 8 de los 45 pacientes): 0.90	Si / No	---

En lo referente a la fiabilidad, la consistencia interna se evaluó en todas las escalas, la reproducibilidad test-retest en todas menos en una¹¹⁴ y la confiabilidad sólo se analizó en cinco escalas^{115-117,120,121}.

Cuando se realizó el test-retest solo hubo seis escalas^{112,113,115,117,119,121} que superaron el número ideal de más de 29 pacientes recomendables para efectuar esta evaluación.

La consistencia interna total de la escala fue excelente para siete cuestionarios^{112-115,118-121}, buena para WUSPI¹²² y aceptable para SST¹¹⁶. Para la adaptación de Membrilla-Mesa et al¹¹⁷ del cuestionario SPADI fue buena para uno de los factores y excelente para el otro.

El tiempo para el test retest empleado para los cuestionarios WORC¹²⁰ y WOSI¹²¹ fue demasiado largo (hasta 21 días) y resultó demasiado corto para el SDQ¹¹⁹ (1 hora). El ICC fue muy bueno en todos los casos.

La confiabilidad fue analizada en cinco cuestionarios^{115-117,120,121}. Membrilla-Mesa et al¹¹⁷ sólo analizaron, en la adaptación al español del cuestionario SPADI el valor del MDC95 que fue de 12'2%. En los otros cuatro cuestionarios se analizaron los valores de EEM y MDC90. Los resultados para ellos fueron, respectivamente para ambos valores, los siguientes: para ULFI de 3'52% y 8'03% (equivalente a 2 puntos de los 25 máximos de la escala)¹¹⁵, para SST de 2'21% y 6'2 %¹¹⁶, para WORC de 192'76 (rango por dominios 39'43-63'82) y 533 (rango por dominios 109'21-176'78) para el rango de 0-2100 de la escala total¹²⁰ y para WOSI de 192'76 (rango por dominios 109'21-176'78) y 533 (rango por dominios 109'21-176'78) para el rango de 0-2100 de la escala total¹²¹.

La validez se evaluó en todos los cuestionarios salvo para las adaptaciones de la escala DASH en Puerto Rico¹¹⁴ y para el SDQ en Méjico¹¹⁹. La validez variaba en función de los instrumentos de comparación, pero, en general, fue adecuada. Para validar las escalas WORC y WOSI en Argentina^{120,121}, se utilizaron, como comparadores, la escala de Constant y una versión adaptada a la población argentina del SST cuyos datos no están publicados en forma de artículo¹²³. Sólo en las adaptaciones de los cuestionarios DASH de Hervás et al¹¹³ y SPADI de Membrilla-Mesa et al¹¹⁷ se evaluó apropiadamente la validez de constructo al haberse formulado a priori una hipótesis relativa a la correlación de los instrumentos que se comparaban.

El análisis de la estructura factorial fue realizado sólo para tres cuestionarios: ULFI¹¹⁵, SST¹¹⁶ y una de las dos adaptaciones del SPADI¹¹⁷, todos ellos adaptados por el mismo grupo de trabajo. En el primer caso la escala resultó unidimensional, en el segundo tridimensional y en el tercero bidimensional. Sólo en dos casos^{115,117} el número de

pacientes utilizado fue apropiado (mas de 100). Los efectos suelo y techo no se analizaron en cinco cuestionarios^{112,113,116,117,119} y, en el resto, solo para el cuestionario WUSPI estaba presente un efecto suelo¹²².

La sensibilidad sólo fue examinada en tres escalas^{113,118}. Para el cuestionario DASH adaptado por Hervás et al¹¹³ se obtuvieron valores para ES y para SRM que representan una sensibilidad grande en ambos casos. Para una de las adaptaciones del SPADI¹¹⁸ los valores fueron, respectivamente, moderados-grandes para la escala global, aunque para la subescala funcional los valores fueron grandes en ambos casos. Los valores de ES y SRM para el cuestionario OSS¹¹⁸ fueron pequeños.

No pudieron analizarse las propiedades clinimétricas de los dos módulos opcionales del DASH, por los pocos pacientes que los contestaron, en la adaptación de Hervás et al¹¹³. Tampoco se analizaron en la versión de Mulero-Portela et al¹¹⁴ para pacientes puertorriqueños.

En cuatro cuestionarios^{114,119,121} las adaptaciones transculturales se realizaron en países latinoamericanos. En tres de los casos¹¹⁹⁻¹²¹ no hay diferencias culturales tan marcadas que impidan el uso del cuestionario en España: la introducción de pequeños cambios en algunas palabras usadas permitiría emplearlos, sin problema, en España. Por ejemplo, en la adaptación para Méjico del SDQ¹¹⁹ se podrían sustituir “auto” por “coche” y “manubrio” por “manillar”. Del mismo modo, en las adaptaciones realizadas en España donde aparecieran palabras malsonantes para sujetos de origen latinoamericano éstas podrían modificarse (sustituir “coger” por “agarrar”) y ser, igualmente, aplicadas allí. En el caso en la adaptación del cuestionario DASH a la población puertorriqueña¹¹⁴ fue necesario permitir algunos anglicismos que no la hacen

aplicable en España. En España pueden usarse cualquiera de las otras dos versiones del DASH^{112,113}.

9.1.2. RODILLA

Mediante la búsqueda bibliográfica realizada en las principales bases de datos biomédicas identificamos diez artículos relevantes¹²⁴⁻¹³³.

Se encontraron diez cuestionarios autocumplimentados que habían sido traducidos y adaptados transculturalmente a la población española y de los que se había estudiado las propiedades clinimétricas de la nueva versión. En todos los casos el cuestionario original se había desarrollado en inglés, excepto en dos, el cuestionario *Arthrose des Membres Inférieurs et Qualité de vie* (AMICAL)¹²⁴ y el *Osteoarthritis Knee and Hip Quality of Life questionnaire* (OAKHQOL)¹²⁹, desarrollados en francés. Había tres cuestionarios aplicables a patologías del miembro inferior: dos para cualquier tipo de afección: *Lower Limb Functional Index* (LLFI)¹³¹, *Lower Extremity Functional Scale* (LEFS)¹³⁰; y uno específico para artrosis: AMICAL¹²⁴. Se encontraron otros tres cuestionarios aplicables a patologías de rodilla y de cadera: *Western Ontario and McMaster Universities Osteoarthritis* (WOMAC) index¹²⁷, OAKHQOL¹²⁹ y *Hip and Knee questionnaire* (HKQ)¹²⁵. Había dos cuestionarios generales de rodilla: *Knee Injury and Osteoarthritis Outcome Score* (KOOS)¹²⁶ y *Knee Society Clinical Rating System* (KSS)¹²⁸. Finalmente se hallaron dos cuestionarios para situaciones específicas de rodilla: *Victorian Institute of Sport Assessment questionnaire* (VISA-P)¹³² para la tendinopatía rotuliana y *Kujala Score*¹³³ para pacientes con dolor femoropatelar. Hay que señalar que nuestra búsqueda no encontró ningún cuestionario desarrollado originalmente en español.

En la tabla 7 se pueden ver las características demográficas y clínicas de la población en la que se realizó cada uno de los artículos.

La escala adaptada, junto con las instrucciones para el paciente, se incluye en la publicación en todos los casos salvo para tres cuestionarios: KOOS¹²⁶, WOMAC¹²⁷ y KSS¹²⁸.

La mayoría de los artículos se han publicado muy recientemente: una de las adaptaciones se publicó en 2015¹³³, cinco en 2014^{124,125,130,131}, una en 2013¹²⁸ y otra en 2012¹²⁶. La primera escala adaptada al español fue el WOMAC, en 2004¹²⁷. Hay que señalar que esta escala se tradujo por primera vez al español en 1999¹³⁴ pero al no incluir el estudio de las propiedades psicométricas fue excluido de nuestra selección de artículos.

Todos los cuestionarios fueron validados en España, excepto uno (AMICAL)¹²⁴ que se validó en Méjico. En todas las adaptaciones, menos en una¹²⁶, se incluyó un número de pacientes superior al mínimo recomendado de 50 para este tipo de estudios.

Tabla 7. Características demográficas y clínicas de la población de los estudios de adaptación transcultural.

ESCALA	Autor (año de la adaptación)	País (Población)	Nº de pacientes incluidos	Diagnóstico	Edad	% de mujeres
AMICAL ¹²⁴	Espinosa – Cuervo et al (2014)	México	100	Gonartrosis grado II-III radiológico de la clasificación de Kellgren & Lawrence	Media: 56,34 (DE±13)	67%
HKQ ¹²⁵	Castellet et al (2014)	España (no especificado)	316	Gonartrosis con indicación de prótesis de rodilla antes y después de la cirugía	≥ 18	----
KOOS ¹²⁶	Vaquero et al (2012)	España (no especificado)	20	Defecto osteocondral y/o defecto osteocondral degenerativo	Media: 41.35 (DE±14) Rango: 24-70	70%
WOMAC ¹²⁷	Villaneuva et al (2004)	España (Sevilla)	73	Gonartrosis (al menos grado II de la clasificación de Kellgren & Lawrence)	Media: 58.71 Rango: 35–73	89%
KSS ¹²⁸	Ares et al (2013)	España (no especificado)	316	Gonartrosis con indicación de prótesis de rodilla antes y a los 6 meses de la cirugía	-----	----
OAKHQOL ¹²⁹	Gonzalez Sáenz de Tejada et al (2011)	España (Canarias)	409	Gonartrosis	Media: 71.26 (DE±7.71)	63.3%
LLFI ¹³¹	Cuesta-Vargas et al (2014)	España (Málaga)	136	Patología crónica (> 12 semanas) de extremidades inferiores	Media: 48 (DE±19)	54.4%
LEFS ¹³⁰	Cruz-Díaz et al (2014)	España (Jaén)	132	Patología de un miembro inferior (un 1.5% diagnosticado de osteoartritis; 22% dolor muscular; 9.1% lesión ligamentosa; 23.5% lesión meniscal; 22,0% dolor femoropatelar, 7.6% fractura. articular, muscular o de partes blandas)	Media: 27.11(DE± 6.22)	44.69%
VISA-P ¹³²	Hernández-Sánchez et al (2011)	España (no especificado)	150	Tendinopatía rotuliana con tendinosis verificada por ecografía o resonancia magnética	Media: 23.5	7.5%
Kujala ¹³³	Gil-Gómez et al (2015)	España (no especificado)	130	Dolor femoropatelar reclutados en clubes deportivos y centros de fisioterapia	Media: 21.2 (DE±3.6)	71.54%

En general el proceso de adaptación transcultural de todas las escalas requirió modificaciones, aunque de poca importancia, respecto a la versión original.

Los principales resultados de la evaluación de la metodología utilizada para la adaptación transcultural se analizan en la tabla 8.

Seis cuestionarios cumplieron rigurosamente los cinco pasos de las recomendaciones de las guías internacionales^{124,126,128,129,132,133}. El resto de cuestionarios fueron menos estrictos en ese aspecto. La prueba piloto se realizó para ocho escalas^{124-129,132,133} aunque ninguno se hizo con más de 30 sujetos.

Tabla 8. Evaluación de la metodología empleada para la adaptación transcultural de los cuestionarios.

ESCALA	Traducción	Síntesis	Traducción inversa	Análisis por comité de expertos	Prueba piloto (número de pacientes)
AMICAL ¹²⁴	+	+	+	+	+ (10)
HKQ ¹²⁵	+	+	-	?	+ (20)
KOOS ¹²⁶	+	+	+	+	+ (10)
WOMAC ¹²⁷	+	+	-	+	+ (10)
KSS ¹²⁸	+	+	+	+	+ (20)
OAKHQOL ¹²⁹	+	+	+	+	+ (17)
LLFI ¹³¹	+	+	-	+	0
LEFS ¹³⁰	+	+	+	+	0
VISA-P ¹³²	+	+	+	+	+ (12)
KUJALA ¹³³	+	+	+	+	+ (19)

Las principales propiedades psicométricas analizadas en las versiones españolas se muestran en la tabla 9.

Ningún trabajo evaluó todas las propiedades métricas de la nueva versión. En general la conclusión de los autores de cada una de las adaptaciones realizadas fue que las propiedades psicométricas evaluadas resultaban aceptables y comparables a las de las versiones originales y a otras versiones adaptadas a otros idiomas.

En lo referente a la fiabilidad, la consistencia interna se evaluó en todas las escalas menos en una¹²⁸, la reproducibilidad test-retest en todas menos en dos^{125,128} y la confiabilidad sólo se analizó en dos escalas^{130,131}. Cuando se realizó el test-retest solo dos escalas^{124,127} indicaron el número de participantes, siendo una¹²⁴ la que superó el número ideal de más de 29 pacientes recomendables para efectuar esta evaluación. La consistencia interna total en todas las escalas, menos en una¹²⁹, fue $> 0.7-0.8$, valor suficiente para garantizar su fiabilidad. El ICC fue muy bueno en todos los casos. La confiabilidad fue analizada en dos cuestionarios^{130,131}. Cruz-Díaz et al¹³⁰ analizaron el valor del MDC95 que fue de 2.18 y el SEM fue de 0.88. Cuesta Vargas et al¹³¹ calcularon el valor de MCD90 y de SEM de 7.12% y 3.12% respectivamente. La validez se evaluó en todos los cuestionarios. Esta variaba en función de los instrumentos de comparación, pero, en general, fue adecuada.

Los efectos suelo y techo sólo se analizaron en cuatro cuestionarios^{125,129,130,133} y en ninguno estaba presente. La sensibilidad sólo fue examinada en dos escalas^{130,132}. En ellos se obtuvieron valores, para ES y SRM, que indicaban una sensibilidad grande en ambos casos.

Tabla 9. Principales propiedades psicométricas analizadas en los cuestionarios adaptados.

ESCALA	Consistencia Interna (α de Cronbach) total	Reproductibilidad Test-retest Número participantes (tiempo entre las dos evaluaciones) ICC	Instrumento de comparación: Validez (coeficiente de correlación) (valores negativos indican correlación inversa)	Efectos suelo / techo	Sensibilidad	
					ES	SRM
AMICAL ¹²⁴	0,946	65 (72h) 0,979 (IC 95%: 0,961-0,991)	- WOMAC: EVA ($r= 0,737$), capacidad funcional ($r= - 0,831$), dolor ($r= -0,742$) y rigidez ($r -0,649$) con $p < 0.05$ - SF-36: actividad física ($r=0,709$); rol físico ($r =0,678$); salud mental ($r= 0,705$); salud general ($r= 0,506$); rol emocional ($r= 0,432$). En otros dominios, las correlaciones fueron bajas pero significativas: dolor ($r= 0,484$), función social ($r= 0,346$) y apoyo social ($r= 0,287$) con $p < 0,05$	-----		---
HKQ ¹²⁵	0,864	No estudiado	- WOMAC: dolor y funcionalidad $r= 0.641$ y rigidez: $r= 0.533$ - SF-36: Funcionalidad ($r=0,458$); rol físico ($r= 0.480$); salud mental ($r=0,333$); salud general ($r=0,316$); rol emocional ($r=0,406$); social ($r=0,405$); vitalidad ($r=0,371$); dolor ($r=0.537$); Componente físico ($r=0,457$); Componente mental ($r=0,360$)	No / No		---
KOOS ¹²⁶	> 0.7 (dolor: 0,88; síntomas: 0,66, AVD: 0,96; deporte/ocio: 0,91, calidad de vida: 0,88)	---- > 0.8	- SF-36: correlación más alta con el dolor ($r >0.7$)	----		---
WOMAC ¹²⁷	Rango de 0,71 a 0,97 (dolor) y de 0,64 a 0,95 (función)	17 (72h) 0,36 a 0,76 para dolor y 0,34 a 0.52 para función	- WOMAC: dolor ($r= 0,27$); función ($r= 0,77$)	-----		-----
KSS ¹²⁸	No estudiado	No estudiado	- WOMAC: significativa con $p < 0.01$ en todos los coeficientes. - SF-36: correlación algo inferior, especialmente dentro de las dimensiones mentales, siendo cerca de 0,4 el resto de las dimensiones.	No / No		----

Tabla 9 (continuación). Principales propiedades psicométricas analizadas en los cuestionarios adaptados.

ESCALA	Consistencia Interna (α de Cronbach) total	Reproductibilidad Test-retest Número participantes (tiempo entre las dos evaluaciones) ICC	Instrumento de comparación: Validez (coeficiente de correlación) (valores negativos indican correlación inversa)	Efectos suelo / techo	Sensibilidad ES SRM
OAKHQOL ¹²⁹	Rango de 0.60 a 0.93	---- Rango de 0.75 a 0.81	- SF-36: moderadamente correlacionado con los dominios físicos (función física, $r = 0,58$; dolor, $r = 0,64$; salud general, $r = 0,41$; y PCS, $r = 0,59$) - WOMAC: moderadamente correlacionado con la función $r = -0.80$ y con el dolor $r = -0.69$ - EDQ-5D: $r > 0.5$	----	----
LLFI ¹³¹	0.91	---- 0.96 (Rango de 0.93 a 0.97)	- WOMAC: fuerte ($r = 0,77$) - EQ-5D-3L: justo e inversamente correlacionados ($r = -0.62$)	---	---
LEFS ¹³⁰	0.98	---- 0,998 (IC 95%: 0,996-0,999)	- SF-36: Funcionalidad ($r = 0,896$); Rol Físico ($r = 0.800$); Rol emocional ($r = 0,504$); Social ($r = 0,764$); Salud mental ($r = 0,744$); Vitalidad ($r = 0,760$); Dolor ($r = 0.903$); Salud general ($r = 0,657$); Componente físico ($r = 0,902$); Componente mental ($r = 0,824$) con $p < 0.01$	No / No	2.3 1.9
VISA-P ¹³²	0.885	---- 0.994 (IC 95%: 0.992-0.996)	- Kujala: $r = 0.897$; $p < 0.001$ - Cincinnati: $r = 0.782$, $p < 0.001$ - SF-36: $r > 0.6$, $p < 0.001$	----	1.14 1.17
Kujala ¹³³	0.8	---- 0.99	- VISA-P: $r = 0,7$; $p < 0,001$	No / No	----

9.1.3. CADERA

Se identificaron ocho artículos relevantes^{124,125,127,129-131,135,136}, que describían ocho cuestionarios autocumplimentados adaptados transculturalmente a la población española y de los que se habían estudiado las propiedades clinimétricas de la nueva versión. Había tres cuestionarios aplicables a cualquier patología del miembro inferior: LLFI¹³¹, LEFS¹³⁰ y el cuestionario AMICAL¹²⁴. Tres cuestionarios eran aplicables a patologías de rodilla y de cadera: WOMAC¹²⁷; OAKHQOL¹²⁹ y HKQ¹²⁵. Había, por último, dos cuestionarios específicos de cadera: *Hip Outcome Score (HOS)*¹³⁵ y el *International Hip Outcome Tool-33 (iHOT-33)*¹³⁶.

En todos los casos el cuestionario original se había desarrollado en inglés, excepto en dos: AMICAL¹²⁴ y OAKHQOL¹²⁹, diseñados en francés.

Nuestra búsqueda no localizó ningún cuestionario desarrollado originalmente en español.

En la tabla 10 se pueden ver las características demográficas y clínicas de la población en la que se realizó cada uno de los artículos.

La escala adaptada, junto con las instrucciones para el paciente, se incluye en la publicación en todos los casos salvo para dos cuestionarios: WOMAC¹²⁷ e iHOT-33¹³⁶. La mayoría de los artículos se han publicado muy recientemente. Una de las adaptaciones se publicó en 2015¹³⁶ y cinco en 2014^{124,125,130,131,135}. La primera escala adaptada al español fue el WOMAC, en 2004¹²⁷. Esta escala se adaptó por primera vez al español en 1999¹³⁴ pero al no incluir el estudio de las propiedades psicométricas fue excluida de nuestra selección de artículos.

Todos los cuestionarios se validaron en España, excepto uno¹²⁴ que se validó en Méjico. En todas las adaptaciones se incluyó un número de pacientes superior al mínimo recomendado de 50 para este tipo de estudios.

Tabla 10. Características demográficas y clínicas de la población de los estudios de adaptación transcultural.

ESCALA	Autor (año de la adaptación)	País (Población)	Nº de pacientes incluidos	Diagnóstico	Edad	% de mujeres
LLFI ¹³¹	Cuesta-Vargas et al (2014)	España (Málaga)	136	Patología crónica (> 12 semanas) de extremidades inferiores	Media: 48 (DE±19)	54,4%
LEFS ¹³⁰	Cruz-Díaz et al (2014)	España (Jaén)	132	Patología de un miembro inferior (un 1,5% osteoartritis; 22% dolor muscular; 9,1% lesión ligamentosa; 23,5% lesión meniscal; 22,0% dolor femoropatelar, 7,6% fractura articular, muscular o de partes blandas)	Media: 27,11 (DE±6,22)	44,7%
AMICAL ¹²⁴	Espinosa – Cuervo et al (2014)	Méjico	100	Coxartrosis grado II-III radiológico clasificación de Kellgren & Lawrence	Media: 56,34 (DE±13)	67%
WOMAC ¹²⁷	Villaneuva et al (2004)	España (Sevilla)	73	Coxartrosis (al menos grado II de la clasificación de Kellgren & Lawrence)	Media: 58,71 Rango: 35–73	89%
OAKHQOL ¹²⁹	Gonzalez Sáenz de Tejada et al (2011)	España (Canarias)	409	Coxartrosis	Media: 71,26 (DE±7,71)	63,3%
HKQ ¹²⁵	Castellet et al (2014)	España (¿?)	316	Coxartrosis con indicación de prótesis de rodilla antes y después de la cirugía	≥ 18	----
HOS ¹³⁵	Seijas et al (2014)	España (¿?)	100	Pinzamiento acetabular mixto (37), pinzamiento tipo Cam (26), Cam +Labrum (15), Tönnis II (5), Tönnis I (5), pinzamiento tipo Pincer (4), lesión del labrum (3), trocanteritis (2), secuelas de epifisiolisis (1), secuela de Perthes (1) y osteonecrosis (1)	Media: 45,1 (DE±12,1) Rango: 18–65	36%

Tabla 10 (continuación). Características demográficas y clínicas de la población de los estudios de adaptación transcultural.

ESCALA	Autor (año de la adaptación)	País (Población)	Nº de pacientes incluidos	Diagnóstico	Edad	% de mujeres
IHOT-33 ¹³⁶	Ruiz-Ibán et al (2015)	España (¿?)	97	Pinzamiento acetabular (78:65 mixto, 11 tipo Cam, 2 tipo Pincer), artrosis(10), patología glúteo medio(3), secuelas de Perthes(2), secuelas de epifisiolisis(2), tendinitis del psoas(1), displasia congénita de cadera(1) y osteonecrosis(1)	Media: 43,8 (DE±10,9) Rango: 22-60	38,1%

En general, el proceso de adaptación transcultural de todas las escalas requirió modificaciones, aunque de poca importancia, respecto a la versión original.

Los principales resultados de la evaluación de la metodología utilizada para la adaptación transcultural se analizan en la tabla 11.

Cuatro cuestionarios cumplieron rigurosamente los cinco pasos de las recomendaciones de las guías internacionales^{124,129,135,136}. El resto fueron menos estrictos en ese aspecto. La prueba piloto se realizó para seis escalas^{124,125,127,129,135,136}, aunque sólo dos^{135,136} la hicieron con 30 pacientes.

Tabla 11. Evaluación de la metodología empleada para la adaptación transcultural de los cuestionarios.

ESCALA	Traducción	Síntesis	Traducción inversa	Análisis por comité de expertos	Prueba piloto (número de pacientes)
LLFI ¹³¹	+	+	-	+	0
LEFS ¹³⁰	+	+	+	+	0
AMICAL ¹²⁴	+	+	+	+	+ (10)
WOMAC ¹²⁷	+	+	-	+	+ (10)
OAKHQOL ¹²⁹	+	+	+	+	+ (17)
HKQ ¹²⁵	+	+	-	?	+ (20)
HOS ¹³⁵	+	+	+	+	+ (30)
IHOT-33 ¹³⁶	+	+	+	+	+ (30)

Las principales propiedades psicométricas analizadas en las versiones españolas se muestran en la tabla 12.

Ningún trabajo evaluó todas las propiedades métricas de la nueva versión. En general, la conclusión de los autores de cada una de las adaptaciones realizadas fue que las propiedades psicométricas evaluadas resultaban aceptables y comparables a las de las versiones originales y a otras versiones adaptadas a otros idiomas.

En lo referente a la fiabilidad, la consistencia interna se evaluó en todas las escalas; en la reproducibilidad test-retest, sólo se especifica en cuatro de ellas el tiempo entre escalas^{124,127,135,136} y de esas, sólo en dos^{124,127} el número de pacientes a los que se

aplicó, siendo una¹²⁴ la que superó el número ideal de más de 29 pacientes recomendables para efectuar esta evaluación.

La consistencia interna total en todas las escalas fue superior a 0,7-0,8, valor suficiente para garantizar su fiabilidad. El ICC fue muy bueno en todos los casos.

El grado de acuerdo fue analizado en dos cuestionarios^{135,136}. Seijas et al¹³⁵ analizaron el valor de MCD, que fue de 13,7 para la subescala de actividades de la vida diaria (AVD) y de 22,8 para la subescala de deporte, y el EEM fue de $\pm 5,1$ en AVD y $\pm 8,5$ para deporte. Ruiz-Ibán et al¹³⁶ calcularon el valor de MCD y de EEM de 12,5 y 4,66 respectivamente.

La validez se evaluó en todos los cuestionarios. Esta variaba en función de los instrumentos de comparación, pero, en general, fue adecuada.

Los efectos suelo y techo sólo se analizaron en cuatro cuestionarios^{125,130,135,136}, estando presente el efecto techo en uno¹³⁵ y el efecto suelo en otro¹³⁶.

La sensibilidad se examinó sólo en tres escalas^{130,135,136}. En dos de ellas sólo se evaluó el RME^{135,136}, y sólo en una¹³⁰ se obtuvieron valores, para TE y RME, que indicaban una sensibilidad grande en ambos casos.

Tabla 12. Principales propiedades psicométricas analizadas en los cuestionarios adaptados.

ESCALA	Consistencia Interna (α de Cronbach) total	Reproducibilidad Test-retest: Número participantes (tiempo entre las dos evaluaciones) ICC	Instrumento de comparación: Validez (coeficiente de correlación) (valores negativos indican correlación inversa)	Efectos suelo / techo	Sensibilidad TE RME
LLFI ¹³¹	0,91	---- 0,96 (Rango de 0,93 a 0,97)	- WOMAC: fuerte ($r= 0,77$) - EQ-5D-3L: justo e inversamente correlacionados ($r = -0,62$)	---	---
LEFS ¹³⁰	0,98	---- 0,998 (IC 95%: 0,996-0,999)	- SF-36: Funcionalidad ($r=0,896$); Rol Físico ($r= 0,800$); Rol emocional ($r=0,504$); Social ($r=0,764$); Salud mental ($r=0,744$); Vitalidad ($r=0,760$); Dolor ($r=0,903$); Salud general ($r=0,657$); Componente físico ($r=0,902$); Componente mental ($r=0,824$), $p<0,01$	No / No	2,3 1,9
AMICAL ¹²⁴	0,946	65 (72h) 0,979 (IC 95%: 0,961-0,991)	- WOMAC: EVA ($r= 0,737$), capacidad funcional ($r= -0,831$), dolor ($r= -0,742$) y rigidez ($r -0,649$) con $p< 0,05$ - SF-36: actividad física ($r=0,709$); rol físico ($r =0,678$); salud mental ($r= 0,705$); salud general ($r= 0,506$); rol emocional ($r= 0,432$). En otros dominios, las correlaciones fueron bajas pero significativas: dolor ($r= 0,484$), función social ($r= 0,346$) y apoyo social ($r= 0,287$) con $p < 0,05$	-----	---
WOMAC ¹²⁷	Subescala: - Dolor: 0,71 a 0,97 - Función: 0,64 a 0,95	17 (72h) 0,36 a 0,76 para dolor y 0,34 a 0,52 para función	- WOMAC: dolor ($r= 0,27$); función ($r= 0,77$)	-----	-----
OAKHQOL ¹²⁹	Rango de 0,60 a 0,93	---- Rango de 0,75 a 0,81	- SF-36: moderadamente correlacionado con los dominios físicos (función física, $r = 0,58$; dolor, $r = 0,64$; salud general, $r = 0,41$; y PCS, $r = 0,59$) -WOMAC: moderadamente correlacionado con la función $r = -0,80$ y el dolor $r=-0,69$ - EDQ-5D: $r > 0,5$	----	----

Tabla 12 (continuación). Principales propiedades psicométricas analizadas en los cuestionarios adaptados.

ESCALA	Consistencia Interna (α de Cronbach) total	Reproducibilidad Test-retest: Número participantes (tiempo entre las dos evaluaciones) ICC	Instrumento de comparación: Validez (coeficiente de correlación) (valores negativos indican correlación inversa)	Efectos suelo / techo	Sensibilidad TE RME
HKQ ¹²⁵	0,864	No estudiado	- WOMAC: dolor y funcionalidad $r=0,641$ y rigidez: $r=0,533$	No / No	---
HOS ¹³⁵	Subescala: - AVD: 0,95; - Deporte: 0,9	--- (15 días) Subescala: - AVD: 0,95 (IC 95%: 0,92-0,97); - Deporte: 0,94 (IC 95%: 0,89-0,97).	- WOMAC: Subescala AVD: Dolor ($r=0,699$); Rigidez ($r=0,667$); función ($r=0,788$). Subescala deporte: Dolor ($r=0,607$); Rigidez ($r=0,499$); función ($r=0,764$)	No / Sí (para subescala deporte)	Subescala: AVD ----- 1,53 Deporte ----- 1,27
IHOT-33 ¹³⁶	Global: 0,98 Subescala: - Función: 0,97; - Deporte: 0,94 ; - Trabajo: 0,89; - Social: 0,94	--- (15 días) Global: 0,97 (IC 95%: 0,96-0,99) Subescala: - Función: 0,95 (IC 95 %: 0,92-0,98); - Deporte: 0,92 (IC 95 %: 0,76-0,98); - Trabajo: 0,93 (IC 95 %: 0,83-0,98); - Social: 0,96 (IC 95 %: 0,91- 0,98)	- WOMAC: Subescala Funcional: dolor ($r=0,744$); rigidez ($r=0,687$); función ($r=0,79$); Subescala Deporte: dolor ($r=0,526$); rigidez ($r=0,513$); función ($r=0,536$); Subescala Trabajo: dolor ($r=0,581$); rigidez ($r=0,544$); función ($r=0,616$); Subescala Social: dolor ($r=0,653$); rigidez ($r=0,574$); función ($r=0,658$)	Sí / No	---- 1,18

En el cuestionario AMICAL¹²⁴ la adaptación transcultural se realizó en un país hispanoamericano (Méjico). No hay diferencias culturales tan marcadas que dificulten el uso del cuestionario en España: la introducción de pequeños cambios en algunas palabras usadas permitiría emplearlos. Por ejemplo, se podría sustituir la palabra “pesero” por “autobús”.

9.2. SEGUNDA FASE

Al finalizar la primera fase de la reunión de consenso de expertos se seleccionaron cuatro de las nueve escalas totales (DASH, SST, SPADI y OSS).

La escala DASH se descartó por abarcar todo el miembro superior y no ser específica del hombro. Es una escala que requiere un tiempo mínimo de 15 minutos para realizarla por lo que su uso habitual presenta dificultades. No se descartó desde el principio por si pudiera ser útil tener una escala que evaluara el miembro superior.

La segunda escala (SST) se rechazó por ser una escala que evaluaba sólo la funcionalidad y había que asociar una escala de dolor.

El resultado final fue que el grupo 1 (formado por traumatólogos y reumatólogos) eligió la OSS y el grupo 2 (formado por rehabilitadores) la SPADI.

SPADI, es una escala que se tarda 2-3 minutos en pasar y es la más utilizada para la patología del manguito rotador (patología mas frecuente en la consulta). Si se compara con la OSS, la SPADI es la más utilizada según la evidencia científica y la emplean todas las especialidades.

Para pasar la OSS se necesitan 4-5 minutos (algo más de tiempo) y, según la literatura científica se utiliza poco. Es más utilizada en Traumatología, pero no en otras especialidades médicas.

A la vista de las consideraciones reflejadas anteriormente, se decidió que la SPADI sería la escala que se integraría en el sistema informático del hospital para evaluar a los pacientes.

10. DISCUSIÓN

10.1. Escalas de valoración

Las medidas de resultados, realizadas desde la perspectiva del paciente, son importantes para evaluar la eficacia de los tratamientos realizados. Las mediciones efectuadas por el clínico, de hecho, se correlacionan mal con la discapacidad percibida por el paciente y este aspecto es, además, el más relevante para él¹³⁷.

Por ello los cuestionarios autocumplimentados por el propio paciente son herramientas útiles que complementan el examen clínico.

El interés por utilizar instrumentos basados en el paciente en la práctica clínica para la evaluación y el seguimiento del tratamiento de pacientes individuales está creciendo. Estos instrumentos permiten a los médicos detectar y tratar problemas funcionales y psicológicos que de otra manera pueden pasarse por alto. Además, promueven la toma de decisiones compartidas y facilitan la comunicación médico-paciente¹³⁸.

A nivel internacional se han desarrollado numerosos cuestionarios para valorar el impacto autopercebido por pacientes.

Cada cuestionario presenta ventajas y desventajas respecto a los demás y ninguno es superior a otro.

El interés que despierta este asunto es creciente y se han publicado numerosas revisiones en los últimos años¹³⁹⁻¹⁵¹.

La mayoría de los cuestionarios se han desarrollado en países anglosajones.

El español es el segundo idioma mas hablado en el mundo, con una amplia extensión geográfica⁷² tanto en países que lo emplean como lengua oficial como utilizado por inmigrantes que residen en países con otras lenguas maternas.

Es preferible adaptar una escala ya existente y que ha demostrado su eficacia, comprobando que la nueva versión mantiene las propiedades psicométricas de la original, que crear otra nueva. Esto último supone un mayor esfuerzo económico y de tiempo e incrementa la diversidad de cuestionarios.

Una misma escala (adaptada apropiadamente cuando sea necesario) facilita las comparaciones entre diferentes poblaciones.

La adaptación transcultural debe asegurar la equivalencia conceptual con la versión original (ser cultural y lingüísticamente apropiada). Hay que tener en cuenta no sólo la adaptación de las palabras sino también giros idiomáticos, contexto cultural y diferencias en la percepción de la salud y la enfermedad.

El número de revisiones sistemáticas de las propiedades psicométricas de los instrumentos de medición del estado de salud ha aumentado de menos de 5 por año antes de 1996 a 45 en 2008 (www.cosmin.nl). Sin embargo, la metodología de las revisiones sistemáticas de las propiedades de medición aún está en desarrollo.

No hay todavía un consenso internacional claro sobre el modo correcto de realizar la adaptación transcultural¹⁰². Existen más de 30 métodos y no se dispone de un *gold standar*. Sí parece haber acuerdo en que la traducción inversa o retrotraducción y la prueba piloto (pretest) resultan esenciales¹⁰².

Es conocido y asumido por la comunidad científica que la adaptación de test no es meramente una cuestión lingüística, y que exige la conjunción de aspectos culturales, conceptuales, lingüísticos y métricos que han de acometerse desde perspectivas de análisis tanto analítico-rationales como empíricas¹⁵³.

En los últimos años la adaptación de los test de unas culturas a otras se ha incrementado en todos los ámbitos evaluativos. Vivimos en un entorno cada vez más multicultural y multilingüe y en el que los test y cuestionarios asisten diariamente en los ámbitos educativo, social, jurídico o clínico, entre otros, en la toma de decisiones individuales o grupales¹⁵².

Numerosos expertos¹⁵⁴⁻¹⁵⁶ consideran que las nuevas tecnologías están influyendo sobre todos los aspectos de la evaluación de los test, tales como el diseño, la construcción y presentación de los ítems y la puntuación. Todo ello está haciendo cambiar el formato y contenido de los cuestionarios, surgiendo la duda razonable de si los test de papel y lápiz, tal como los conocemos hasta ahora, serán capaces de resistir a este nuevo cambio tecnológico.

El uso internacional de los test está atrayendo la atención de investigadores y profesionales, debido a la globalización creciente y a las facilidades de comunicación, lo que plantea problemas de adaptación de los test de unos países a otros^{157,158}. Esta línea de desarrollo es la que justifica y motiva precisamente parte del trabajo que aquí se presenta.

En las diferentes naciones o poblaciones hispanohablantes pueden existir palabras de uso exclusivo o con significado diferente a los países o grupos culturales. No obstante, en general, no hay diferencias culturales tan marcadas, que impidan el uso de estos cuestionarios en los diferentes países que hablan español.

10.1.1. Escalas de hombro

En cuanto a los cuestionarios de hombro, el cuestionario DASH es el más extendido, junto con su versión simplificada de 11 ítems (QuickDASH)¹⁴⁰, no adaptada al español¹⁵⁹. Es el más adaptado transculturalmente¹⁴²: se ha adaptado a más de 35 idiomas y dialectos¹⁴⁰.

Las propiedades psicométricas de la versión de Hervás et al¹¹³, realizada en España, han sido estudiadas en profundidad, aunque no se analizaron los valores de confiabilidad.

Es un cuestionario largo y no específico del hombro sino para todo el miembro superior por lo que en general podría recomendarse preferentemente para la investigación.

Otra opción sería emplear el cuestionario ULFI cuya adaptación transcultural ha sido menos rigurosa¹¹⁵. Sobre ella, a diferencia de la escala DASH, disponemos de datos de confiabilidad, pero no de sensibilidad.

Las propiedades psicométricas de los cuestionarios SPADI y OSS también se han analizado ampliamente y de forma rigurosa salvo los valores de confiabilidad.

Ambas son muy breves.

El cuestionario OSS es el más usado por los cirujanos ortopédicos en el Reino Unido¹⁶⁰ y es la escala general de hombro más adaptada y validada internacionalmente. Sin embargo, se usa relativamente poco en la literatura científica y no es apropiada para pacientes con inestabilidad.

El cuestionario SPADI es más sensible y muy aceptado también por la comunidad científica¹⁶¹. Sería el más recomendable tanto para la práctica clínica como para la investigación^{140,141} y probablemente en más recomendable en atención primaria¹⁶².

Los cuestionarios SST y SDQ son también breves y aplicables a cualquier patología de hombro.

El primero es muy usado en Estados Unidos de Norteamérica¹⁴⁰ pero como sólo evalúa la funcionalidad debería asociarse a una escala de dolor. Además, las respuestas binarias admiten pocos matices o grados.

El segundo parece tener un efecto techo importante, aspecto no analizado en la versión española, donde tampoco se estudió la validez.

Los cuestionarios WORC, WOSI y WUSPI son específicos, respectivamente, para pacientes con patología del manguito rotador, inestabilidad de hombro y usuarios de silla de ruedas.

Los dos primeros se han adaptado a la población argentina^{120,121}. Ambos resultan largos de aplicar. De hecho, en publicaciones sobre roturas del manguito rotador la escala más usada es el SST¹⁶².

En el tercero¹²² la calidad de la adaptación transcultural y la evaluación de la validez son deficientes, aunque la fiabilidad es buena.

En base a los resultados de nuestra revisión, dentro de los cuestionarios de hombro adaptados al español, consideramos que el SPADI es la primera opción tanto para la práctica clínica como para la investigación sobre el dolor de hombro, y que el DASH es preferible para su uso en investigación del miembro superior en general, ya que no es específico para el hombro.

Entre los pacientes con enfermedad del manguito de los rotadores, el SPADI sería preferible al WORC porque la adaptación de este último nos parece de peor calidad.

Para otras situaciones específicas, como la inestabilidad de hombro y los usuarios de sillas de ruedas, existen los cuestionarios WOSI y WUSPI, respectivamente. Estas adaptaciones son de calidad inferior a la del cuestionario SPADI.

10.1.2. Escalas de rodilla y cadera

El cuestionario WOMAC es el más usado internacionalmente para valorar la capacidad funcional de pacientes con artrosis de rodilla y/o cadera y para evaluar la eficacia de los resultados de tratamientos conservadores y quirúrgicos^{163,164}.

Se compone de 24 ítems y el paciente tarda en contestarlo alrededor de 5-10 minutos¹⁴⁴.

Se han elaborado dos versiones reducidas, con sólo once y catorce cuestiones, con el objetivo de conseguir la misma información que el cuestionario original, pero en menos tiempo. Se ha comprobado que ambas versiones en español conservan las mismas propiedades psicométricas que la versión de 24 ítems¹⁶⁵⁻¹⁶⁷. Estas nuevas versiones se pueden emplear en la práctica clínica y en la investigación.

Como aspectos negativos la escala WOMAC no evalúa aspectos relacionados con la calidad de vida y no permite discriminar cuál es la articulación responsable en caso de que la cadera y la rodilla están afectadas simultáneamente. Sus puntuaciones pueden verse modificadas por la presencia de dolor lumbar, alteraciones en el estado de ánimo y otras patologías extraarticulares¹⁶⁸.

El cuestionario OAKHQOL es muy completo pues incluye aspectos relacionados con salud mental, apoyo social, función social, relaciones personales, actividad sexual y vida profesional, pero el elevado número de ítems que lo componen dificulta su empleo en la práctica clínica¹⁶⁹.

El cuestionario HKQ consta de 16 preguntas. Su validación al español se realizó sólo en artroplastias de rodilla sin incluir pacientes con patologías de cadera. Eso podría representar un inconveniente si se aplica a ese tipo de pacientes.

El cuestionario LEFS incluye 20 preguntas. A diferencia del “WOMAC” permite discriminar entre dolor y función física. Se cumplimenta en 2 minutos y se puntúa en 20 segundos.

El cuestionario LLFI, de 25 ítems, todavía se ha empleado poco debido a su muy reciente creación.

El cuestionario AMICAL es el único específico para valorar la calidad de vida de los pacientes con artrosis de miembros inferiores. Su desventaja es que incluye 43 cuestiones y el paciente tarda en rellenarlo alrededor de 15 minutos¹⁶⁹.

Para la patología general de rodilla, el cuestionario KOOS se elaboró en 1998. Su uso es muy amplio tanto para pacientes con artrosis como traumatismos de la rodilla entre los que se incluyen reparación de ligamentos cruzados y lesiones de meniscos. Además, ha mostrado ser muy efectivo en pacientes jóvenes y activos. Se tarda 10 minutos en cumplimentar¹⁷⁰.

Otra de las escalas de valoración más usadas internacionalmente en la cirugía protésica de rodilla es la KSS. Es una de la más adaptadas transculturalmente. Tiene dos versiones: la original diseñada en 1989 y la modificada en 1993. La original tenía el problema de que la puntuación descendía con el deterioro general del paciente. Por eso se ideó una modificación que separaba los aspectos funcionales de los relacionados con la habilidad del paciente¹²⁸.

Los cuestionarios VISA-P y *Kujala* son escalas para patologías específicas de rodilla (tendinopatía rotuliana y dolor femoropatelar respectivamente). Son muchos los estudios en diferentes países que los han utilizado como una herramienta de evaluación.

Los cuestionarios clásicos diseñados para valorar resultados en pacientes mayores con artrosis de cadera son menos útiles en individuos jóvenes y activos con patología no artrósica y con mayores demandas funcionales.

Disponemos de dos cuestionarios adaptados para estos pacientes. El cuestionario HOS que contiene 28 ítems y evalúa específicamente aspectos funcionales (en un reciente metanálisis¹⁷¹ se recomienda su uso para valorar los resultados tras artroscopia de cadera) y el cuestionario iHOT-33 que valora, a diferencia del HOS, aspectos emocionales, sociales, laborales y del estilo de vida. Existe una versión reducida de 12

preguntas (iHOT-12)¹⁷² que podría ser más útil en la clínica, aunque aún no está validada en español.

Actualmente disponemos de veintiuna escalas para medir la limitación funcional del hombro, rodilla y cadera adaptadas a la población española siguiendo las recomendaciones internacionales (en 13 casos con el máximo rigor). Todas tienen propiedades psicométricas aceptables y similares a las de los cuestionarios originales y a otras versiones adaptadas a otros idiomas.

No hay un consenso claro todavía para evaluar la calidad de los cuestionarios de medición de salud subjetivos.

Los criterios que utilizamos para evaluar la calidad de los cuestionarios pueden ser discutidos, pero no ha sido nuestra intención crear una lista de verificación de evaluación estandarizada, sino proporcionar información acerca de las propiedades clinimétricas de los cuestionarios disponibles para facilitar su elección en la práctica clínica.

10.2. Panel de expertos

En los últimos años se han publicado en las revistas relacionadas con las ciencias de la salud, tanto nacionales como internacionales¹⁷³⁻¹⁷⁵, trabajos sobre la metodología cualitativa, que incluyen entre sus técnicas las llamadas técnicas de consenso.

Dentro de estas técnicas distinguimos diferentes aproximaciones¹⁷⁶. Una de ellas, son las aproximaciones subjetivas, basadas en la opinión de expertos, a partir de valoraciones implícitas de revisores, métodos cualitativos de consenso u otras técnicas grupales^{177,178}.

Estas aproximaciones permiten abordar objetivos diversos, por lo que los utilizan más las agencias gubernamentales que tienen que valorar líneas de investigación en campos muy diferentes, considerar objetivos múltiples y tener en cuenta las reacciones negativas. Por el contrario, sus resultados dependen de los conocimientos, la experiencia, la representatividad y los intereses de los participantes, y su reproductibilidad es mínima¹⁷⁶.

Tomar decisiones en grupo puede resultar complejo debido a la diferencia de conocimientos de sus miembros, a la fuerza de convicción de los argumentos dados o incluso a los intereses particulares⁷⁵.

Es difícil determinar qué se considera un médico experto: es común que el nivel de los conocimientos teóricos sea máximo al inicio de la vida profesional, cuando el médico novel está recién egresado de una facultad, pero nadie le consideraría experto por ello, ni confiaría a la pericia del recién graduado la atención de casos complejos o críticos.

Para devenir en experto, el médico debe además transitar regularmente una vía formativa imprescindible y simultánea a la del mantenimiento actualizado de su saber, como es la de acumular experiencia clínica^{179,180}. Se requiere un proceso en perfeccionamiento continuo que permite configurar un modelo personal, cada vez más completo y sofisticado, de cada enfermedad o síndrome, “enriquecido” con la experiencia, que se empleará como patrón de comparación para afrontar con mayor eficacia nuevos casos.

Otros autores han discutido sobre si este el lento proceso de adquisición de experiencia que requiere la exposición continuada a la tarea clínica podría acelerarse y asegurarse a través del análisis teórico de casos, de la práctica clínica en entornos simulados (mediante software o simulación robótica), mediante procedimientos de *coach* o de otras estrategias educativas^{181,182}.

Existe, pues, un vínculo estrecho entre las necesidades de información del clínico durante el ejercicio y el objetivo de un experto en cualquier otra disciplina cuando intenta hacer previsiones sobre el devenir futuro de un determinado problema en un contexto incierto¹⁸³.

Las recomendaciones basadas en el criterio de los expertos siguen siendo una ayuda estimable para tomar decisiones en la práctica rutinaria, en particular cuando el clínico afronta escenarios de incertidumbre por ausencia de pruebas sólidas, por controversia científica entre distintas pruebas disponibles, o por la limitada posibilidad de aplicar ciertas evidencias en un contexto clínico concreto (ancianidad, comorbilidad, intolerancia medicamentosa, riesgo de interacciones farmacológicas, situaciones de baja adherencia a recomendaciones)¹⁸³.

11. LIMITACIONES

Y

PUNTOS FUERTES DEL ESTUDIO

- **LIMITACIONES**

Aunque se han hecho todos los esfuerzos posibles para identificar y recuperar todos los cuestionarios relevantes ya adaptados al español, debe decirse que la búsqueda de información se ha hecho en las bases de datos más relevantes (incluidas las bases de datos locales) por lo que pueden existir estudios que estén publicados en revistas no indexadas en estas bases de datos.

Una vez dicho esto, destacamos que no se han encontrado publicaciones de revisiones similares sobre cuestionarios para patologías de las articulaciones de hombro, rodilla y cadera en español, lo que ha dificultado la discusión de nuestro trabajo.

Puede existir un sesgo en la selección de los participantes del panel de expertos, dado que todos trabajan en el Hospital Universitario Fundación Alcorcón, por lo que sería conveniente en un futuro reclutar a participantes de distintos centros sanitarios. Los juicios subjetivos de los participantes pueden llevar a una mayor valoración de los tópicos en los que todos coinciden, dejando al margen ideas más novedosas.

- **PUNTOS FUERTES**

En primer lugar, la originalidad del trabajo, dado que no hemos encontrado en la literatura científica ningún trabajo similar.

Los integrantes en el panel de expertos fueron seleccionados por su especialidad médica, capacidad de liderazgo y para definir el objetivo y por componer un grupo relacionado con distintas áreas de conocimiento y sectores de investigación en servicios de salud.

Tras el trabajo realizado, se ha conseguido seleccionar la escala SPADI como cuestionario de hombro más indicado para su utilización en futuros trabajos de investigación en esta patología.

Gracias a los conocimientos adquiridos y los ítems encontrados en el camino a lo largo de estos cuatro años, hemos conseguido una extensa distribución de los procesos mediante la publicación de tres artículos relacionados con este trabajo (*anexos 2, 3 y 4*).

12. CONCLUSIONES

1. Actualmente disponemos de veintiuna escalas autocumplimentadas para medir la limitación funcional del hombro, rodilla y cadera adaptadas a la población española.
2. De estas veintiuna escalas, doce se han publicado en los últimos cuatro años.
3. Las veintiuna escalas tienen propiedades psicométricas aceptables y similares a las de los cuestionarios originales y a otras versiones adaptadas a otros idiomas.
4. Las escalas que se centran en la valoración funcional son susceptibles de utilizarse en la práctica clínica habitual.
5. De las escalas de hombro adaptadas al español, consideramos que la SPADI es la más indicada para su uso en la práctica clínica, teniendo en cuenta los criterios metodológicos (propiedades psicométricas) y de aplicabilidad (tiempo necesario para cumplimentarla, utilidad para determinadas patologías).

13. BIBLIOGRAFÍA

1. García-Martínez F, Herrera-Silva J, Aguilar-Luque J. Tratamiento del dolor crónico en Atención Primaria. *Rev Soc Esp Dolor*. 2000;7:453-9.
2. IASP Pain terms: a current list with definitions and notes on usage. *Pain*. 1986;3:215-21.
3. Breivik H, Collett B, Ventafridda V, Cohen R, Gallacher D. Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment. *Eur J Pain*. 2006;10:287-333.
4. Loza E, Lopez-Gomez JM, Abasolo L, et al. Economic burden of knee and hip osteoarthritis in Spain. *Arthritis Rheum*. 2009;61:158-65.
5. Briggs AM, Cross MJ, Hoy DG, Sánchez-Riera L, Blyth FM, Woolf AD, March L. Musculoskeletal Health Conditions Represent a Global Threat to Healthy Aging: A Report for the 2015 World Health Organization World Report on Ageing and Health. *Gerontologist*. 2016;56: 243-55.
6. Altman RD. Early management of osteoarthritis. *Am J Manag Care*. 2010;16:41-7.
7. Urwin M, Symmons D, Allison T, Brammah T, Busby H, Roxby M, Simmons A, Williams G. Estimating the burden of musculoskeletal disorders in the community: the comparative prevalence of symptoms at different anatomical sites, and the relation to social deprivation. *Ann Rheum Dis*. 1998;57:649–55.
8. Frau-Escales P, Langa-Revert Y, Querol-Fuentes F, Mora-Amérigo E, Such-Sanz A. Shoulder musculoskeletal disorders in primary care. A cross-sectional study in a health care center of the Valencia Health Care Agency. *Fisioterapia*. 2013;35:10-17.
9. Carmona-Terés V, Lumillo-Gutiérrez I, Jodar-Fernández L, Rodríguez-Blanco T, Moix-Queraltó J, Pujol-Ribera E, Mas X, Batlle-Gualda E, Gobbo-Montoya M, Berenguera A1. Effectiveness and cost-effectiveness of a health coaching

- intervention to improve the lifestyle of patients with knee osteoarthritis: cluster randomized clinical trial. *BMC Musculoskelet Disord.* 2015;16:38.
10. Klintberg IH, Cools AM, Holmgren TM, Holzhausen AG, Johansson K, Maenhout AG, et al. Consensus for physiotherapy for shoulder pain. *Int Orthop.* 2014.
 11. Arnalich Jiménez MB, Sánchez Parera R. Guías Clínicas en Atención Primaria. <http://www.fisterra.com/>. Hombro doloroso. *Guías Clínicas.* 2003;3(10).
 12. Mitchell C, Adebajo A, Hay E, Carr A. Shoulder pain: diagnosis and management in general practice. *BMJ.* 2005;33:1124-8.
 13. Barr KP. Rotator cuff disease. *Phys Med Rehabil Clin North Am.* 2004;15(2):475-91.
 14. Luime JJ et al. Prevalence and incidence of shoulder pain in the general population; a systematic review. *Scand J Rheumatol.* 2004;33:73-81.
 15. Gómez MR, Diz PG, Barros GL, Gómez CS, Lozano Eire MJ, Robisco LP, et al. Estudio epidemiológico de la patología dolorosa del hombro en nuestro medio. *Rev Esp Reumatol.* 1997;24:247-50.
 16. Gómez Raso N, López Martínez N. Patología del manguito de rotadores. ¿Un cajón de sastre? *Rev Soc Madr Med Fam Comunitaria.* 2001;1:18-20.
 17. Vicente Pardo JM. Hombro doloroso e incapacidad temporal. El retorno al trabajo tras larga baja por hombro doloroso. Causalidad del trabajo en el hombro doloroso. *Med Segur Trab.* 2016;62:337-59.
 18. Speed C. Shoulder pain. *BMJ Clin Evid.* 2008;01:1107.
 19. Calmbach, W. Evaluation of patients presenting with knee pain: part I. History, physical examination, radiographs, and laboratory tests. *American family physician.* 2003;68:5.

20. Jackson JL, O'Malley PG y Kroenke K. Evaluación del Dolor de Rodilla Agudo en la Atención Primaria. *Annals of Internal Medicine*. 2003;139:575-88.
21. Guccione AA, Felson DT, Anderson JJ, et al. The effects of specific medical conditions on the functional limitations of elders in the Framingham study. *Am J Publ Health* 1994;84:351-8.
22. Murray CJL, López AD. *The global burden of disease*. Ginebra: World Health Organisation; 1997.
23. Pérez Martín A. Symptoms. Localizations: knee, hip, hands, spine, other localizations. *Aten Primaria*. 2014;46:11-7.
24. Millán Soria J, Fuertes Fortea A, López Grima ML. "Exploración de la cadera en atención primaria. *Revista Valenciana de Medicina de familia*, nº18.
25. Croft P, Pope D, Silman A. The clinical course of shoulder pain: prospective cohort study in primary care. *Br Med J*. 1996;313:601-2.
26. Robinson C, Fatone S. You've heard about outcome measures, so how do you use them? Integrating clinically relevant outcome measures in orthotic management of stroke <http://journals.sagepub.com/doi/10.1177/0309364612446650>.
27. Pérez de Heredia Torres M, Brea Rivero M, Martínez Piédrola RM, Máximo Bocanegra N, Sánchez Camarero C, Calvo Til MJ, Miangolarra Page JC. Clasificación Internacional de Funcionamiento (CIF) y Terapia Ocupacional: ¿un mismo paradigma?. *World Federation of Occupational Therapists Bulletin*. 2016;47:37-43.
28. Ayuso-Mateos JL, Nieto-Moreno M, Sánchez-Moreno J, Vázquez-Barquero JL. Clasificación Internacional del Funcionamiento, la Discapacidad y la Salud (CIF): aplicabilidad y utilidad en la práctica clínica.

- http://sid.usal.es/idocs/F8/ART13069/clasificacion_internacional_del_funcionamiento.pdf
29. Sección de Medicina Física y Board Europeo de Medicina Física y Rehabilitación, Unión Europea de Médicos Especialistas (UEMS) y Academia Europea de Medicina de Rehabilitación. Libro blanco de Medicina Física y Rehabilitación en Europa. Publicación europea de Medicina Física y Rehabilitación; 2009. p.12. Vol. 45.
30. Ilieva EM, Oral A, Küçükdeveci AA, Varela E, Valero R, Berceanu M, Christodoulou N. Osteoarthritis. The role of Physical and Rehabilitation Medicine Physicians. The European perspective based on the best evidence. A paper by the UEMS-PRM Section Professional Practice Committee. Eur J Phys Rehabil Med. 2013;49(4):579-93.
31. https://www.icf-research-branch.org/images/ICF%20Core%20Sets%20Download/Comprehensive_and_Brief_ICF_Core_Sets_Osteoarthritis.pdf
32. Libro Blanco de Medicina Física y Rehabilitación en Europa. 2009. <http://www.whitebookprm.eu/wp-content/uploads/2017/01/SPANISH-VERSION-PRM-WHITE-BOOK-.pdf>
33. Neumann V, Gutenbrunner C, Fialka-Moser V, Christodoulou N, Varela E, Giustini A, Delarque A. Interdisciplinary team working in physical and rehabilitation medicine. J Rehabil Med. 2010;42(1):4-8.
34. Zhang W, Nuki G, Moskowitz R, Abranson S, Altman R, Arden N et al. OARSI recommendations for the management of hip and knee osteoarthritis, Part II: OARSI evidence-based, expert consensus guidelines. Osteoarthritis Cartilage. 2008;16(2):137-62.

35. Jordan KM, Arden NK, Doherty M, Bannwarth B, Bijlsma JW, Dieppe P, Gunther K, Hauselmann H, et al. EULAR Recommendations 2003: an evidence based approach to the management of knee osteoarthritis: Report of a Task Force of the Standing Committee for International Clinical Studies Including Therapeutic Trials (ESCISIT). *Ann Rheum Dis.* 2003;62(12):1145-55.
36. Hochberg MC, Altman RD, April KT, Benkhalti M, Guyatt G, McGowan J, Towheed T, Welch V, Wells G, Tugwell P; American College of Rheumatology. American College of Rheumatology 2012 recommendations for the use of nonpharmacologic and pharmacologic therapies in osteoarthritis of the hand, hip, and knee. *Arthritis Care Res (Hoboken).* 2012;64(4):465-74.
37. Loew L, Brosseau L, Wells GA, Tugwell P, Kenny GP, Reid R, et al. Ottawa panel evidence-based clinical practice guidelines for aerobic walking programs in the management of osteoarthritis. *Arch Phys Med Rehabil.* 2012;93:1269-85.
38. American Academy of Orthopaedic Surgeons. *Treatment of Osteoarthritis of the Knee (2nd ed)*, American Academy of Orthopaedic Surgeons, Rosemont, IL. 2013.
39. Zhang W, Doherty M, Arden N, Bannwarth B, Bijlsma J, Gunther KP, et al. EULAR evidence based recommendations for the management of hip osteoarthritis: report of a task force of the EULAR Standing Committee for International Clinical Studies Including Therapeutics (ESCISIT) *Ann Rheum Dis.* 2005;64:669-81.
40. American Academy of Orthopaedic Surgeons. *Treatment of Osteoarthritis of the Knee (2nd ed)*, American Academy of Orthopaedic Surgeons, Rosemont, IL. 2013.
41. Mazieres B, Thevenon A, Coudeyre E, Chevalier X, Revel M, Rannou F. Adherence to, and results of, physical therapy programs in patients with hip or knee osteoarthritis. Development of French clinical practice guidelines. *Joint Bone Spine.* 2008;75:589-96.

42. Varela E, Valero R, Küçükdeveci AA, Oral A, Ilieva E, Berceanu M, Christodoulou N; UEMS-PRM Section Professional Practice Committee. Shoulder pain management. The role of physical and rehabilitation medicine physicians. The European perspective based on the best evidence. A paper by the UEMS-PRM Section Professional Practice Committee. *Eur J Phys Rehabil Med.* 2013;49(5):743-51.
43. González Montalvo JI, Alarcón Alarcón T. Calidad de los instrumentos de valoración funcional en geriatría: del invento de la rueda a la era electrónica. *Rev Esp Geriatr Gerontol.* 2008;43:265-7.
44. Rodríguez Gázquez M.A, Lopera Jaramillo J. Validación de escalas en Salud mental. *Revista CES MEDICINA Volumen 16 No.3 Octubre - Diciembre / 2002.*
45. Abenoza Guardiola, M. Escalas de valoración funcional, instrumentos para medir nuestro producto sanitario. *Rehabilitación.* Vol. 34. Núm. 1. Enero 2000.
46. Quintana Luque. Escalas de valoración funcional del niño con discapacidad. *Rehabilitación Infantil.* Panamericana 2012.
47. Johnston M, Wilkerson D, Maney M. Evaluation of the quality and outcomes of medical rehabilitation programs. DeLisa JA, Gans M, editors. *Physical Medicine and Rehabilitation. Principles and Practice.* 2nd ed. Philadelphia; 1993. P.240-68.
48. García Bascones M, Beltrán Recio C, Mozo Muriel A. Escalas de valoración infantil. *Guía de rehabilitación infantil Madrid. Médica Panamericana;* 2010; 301-7.
49. Muñiz J, Elosua P, Hambleton RK. Directrices para la traducción y adaptación de los tests: segunda edición. *Psicothema.* 2013;25(2):151-7.
50. Epstein J, Santo RM, Guillemin F. A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus. *J Clin Epidemiol.* 2015;68(4):435-41.

51. Maneesriwongul W, Dixon JK. Instrument translation process: a methods review. *J Adv Nurs*. 2004;48(2):175-86.
52. Carvajal A, Centeno C, Watson R, Martínez M, Rubiales A. How is an instrument for measuring health to be validated? *Anales del sistema sanitario de Navarra*. 2011;34(1):63-72.
53. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976)*. 2000;25(24):3186-91.
54. García de Yébenes Proues MA, Rodríguez Salvanés F, Carmona Ortells L. Validation of questionnaires. *Reumatología Clínica*. 1998;110(14):550-6.
55. Anastasi A, Urbina S. *Test psicológicos*. México: Prencice Hall-Hispanoamericana; 1998.
56. Fortin M, Hungler B. *Nursing research: principles and methods*. Philadelphia;1999.
57. Carvajal A, Centeno C, Watson R, Martínez M, Rubiales A. How is an instrument for measuring health to be validated? *Anales del sistema sanitario de Navarra*. 2001;34:63-72.
58. Darr N, Franjoine M, Young B. Pediatric balance scale performance in children who are developing typically and in children with mild developmental delays. *Pediatr Phys Ther*. 2009;21:89-90.
59. Sánchez R, Gómez C. Conceptos básicos sobre la validación de escalas. *Revista Colombiana de Psiquiatría*. 1998;27:121-30.
60. Johnston M, Wilkerson D, Maney M. Evaluation of the quality and outcomes of medical rehabilitation programs. DeLisa JA, Gans M, editors. *Physical Medicine and Rehabilitation. Principles and Practice*. 2nd ed. Philadelphia;1993;240-68.

61. Guyat G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *Journal of chronic diseases*. 1987;40:171-78.
62. Kartz-Leurer M, Rotem H, Lewitus H, Keren O, Meyer S. Functional balance tests for children with traumatic brain injury: within-session reability. *Pediatr Phys Ther*. 2008;20:254-58.
63. Guyatt G, Walter S, Norman G. Measuring change over time: assesing the usefulness of evaluative instruments. *J Chronic Dis*. 1987;40:171-78.
64. Katz-Leurer M, Rotem H, Lewitus H, Keren O, Meyer S. Relationship between balance abilities and gait characteristics in children with mild developmental delays. *Pediatr Phys Ther*. 2009;21:89-90.
65. Ramada-Rodilla JM, Serra-Pujadas C, Delclós-Clanchet GL. Cross-cultural adaptation and health questionnaires validation: Revision and methodological recommendations. *Salud Pública de México*. 2013;55:57-66.
66. Casas Anguita J, Repullo Labrador JR, Pereira Candel J. Medidas de calidad de vida relacionada con la salud. Conceptos básicos, construcción y adaptación cultural. *Medicina Clínica*. 2001;116:789-96.
67. Held SL, Kott KM, Young BL. Standardized Walking Obstacle Course (SWOC): reliability and validity of new functional measurement tool for children. *Pediatr Phys Ther*. 2006;18:23-30.
68. Gualberto Buena-Casal J, Sierra C, Catarro Dios H, De los Santos Roig, M. Situación actual de la evaluación psicológica en lengua castellana. *Papeles del psicólogo*. 2002.
69. Wright FV, Ryan J, Brewer K. Reliability of the Community Balance and Mobility Scale (CB&M) in high-functioning school-aged children and adolescents who have an acquired brain injury. *Brain Inj*. 2010;24:1585-94.

70. Cardoso Ribero C, Gómez-Conesa A, Hidalgo Montesinos MD. Metodología para la adaptación de instrumentos de evaluación. *Fisioterapia*. 2010;32:264-70.
71. Hambleton R. Translating achievement test for use in cross-national studies. *Eur J Psychol Assess*. 1993;9:57-68.
72. Instituto Cervantes. Informe: El español una lengua viva. 2015 (acceso el 6-2-2016)
Disponible en:
<http://www.cervantes.es/imagenes/File/prensa/El%20espaol%20una%20lengua%20viva.pdf>
73. Badía X. Transcultural measurements of quality of life in relation to health, for adaptation by Spain. *Med Clin*. 1995;105(2):56-8.
74. Puga VO, Lopes AD, Costa LO. Assessment of cross-cultural adaptations and measurement properties of self-report outcome measures relevant to shoulder disability in Portuguese: a systematic review. *Rev Bras Fisioter*. 2012;16(2):85-93.
75. Murphy DR, López M. Neck and back pain specific outcome assessment questionnaires in the Spanish language: a systematic literature review. *Spine J*. 2013;13(11):1667-74.
76. Asua Batarrita J. Entre el consenso y la evidencia. *Gaceta Sanitaria*. 2005;19(1):65-70.
77. Berra A, Marín I, Álvarez R. Metodología de Expertos. Consenso en Medicina. Serie Monografías no 14. Granada: Escuela Andaluza de Salud Pública; 1996.
78. Yañez R, Cuadra R. La técnica Delphi y la investigación en los servicios de salud. *Ciencia y Enfermería*. 2008;14(1):9-15.
79. Lorge I, Fox D, Davitz J, Brenner M. A survey of studies contrasting the quality of group performance and individual performance (1920-1957). *Psychological Bulletin*. 1958;55:337-372.

80. Miner FC, Jr. Group versus individual decision making: A investigation of performance measures, decision strategies, and process losses/gains. *Organizational Behaviour and Human Performance*. 1984;33:112-24.
81. Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CFB, Askham J, et al. Consensus development methods, and their use in clinical guideline development. *Health Technology Assessment*. 1998; 2(3).
82. Dalkey NC, Helmer O. An experimental application of the Delphi method to the use of experts. *Management Science*. 1963;9(3):458-467.
83. Pill J. The Delphi Method: substance, context, a critique and an annotated bibliography. *Soc Econ Plan Sci*. 1971;3:37-71.
84. Delbecq AL, Van de Ven AH. A group process model for problem identification and program planning. *Journal of Applied Behavioral Science*. 1971;7(4):466- 492.
85. U.S. Department of Health & Human Services. National Institutes of Health. NIH Consensus Development Program [Internet]. Acceso 4 de agosto de 2013. Disponible en: <http://consensus.nih.gov/>.
86. Peiró S, Portella E. No todo es acuerdo en el consenso. Limitaciones de los métodos de consenso en los servicios de salud. *Gaceta Sanitaria*. 1993;7:294- 300.
87. Astigarraga E. El método Delphi. San Sebastián: Universidad de Deusto; 2003.
88. Williams PL, Webb C. The Delphi technique: a methodological discussion. *J Adv Nurs*. 1994;19(1):180-6.
89. Powell C. The Delphi technique: myths and realities. *J Adv Nurs*. 2003;41(4):376-82.
90. Green B, Jones M, Hughes D, Williams A. Applying the Delphi technique in a study of GPs' information requirements. *Health Soc Care Community*. 1999;7(3):198-205.

91. Dobbie A, Rhodes M, Tysinger JW, Freeman J. Using a modified nominal group technique as a curriculum evaluation tool. *Fam Med.* 2004;36(6): 402-6.
92. Isaacs D, Fitzgerald D. Seven alternatives to evidence based medicine. *BMJ* 1999;319:1618.
93. Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CF, Askham J, et al. Consensus development methods, and their use in clinical guideline development. *Health Technol Assess.* 1998;2:1-88.
94. Jones JH. Consensus methods for medical and health services research. En: Pope CM, Mays N, Editors. *Qualitative research in health care.* Massachusetts: BMJ books. 2000. p. 46.58.
95. Briones E, Marín I, Álvarez R, Reyes A. Fundamentos de consenso en el ámbito de las ciencias de la salud. En: *Consenso en Medicina. Metodología de Expertos.* Granada: Monografías EASP; 1996.
96. Bernal-Delgado E, Peiró S, Sotoca R. Prioridades de investigación en servicios sanitarios en el Sistema Nacional de Salud. Una aproximación por consenso de expertos. *Gac Sanit.* 2006;20:287-94.
97. Pérez Andrés C. ¿Deben estar las técnicas de consenso incluidas entre las técnicas de investigación cualitativa?. *Rev Esp Salud Pública.* 2000;74:319-21.
98. Olaz AJ. La técnica del grupo nominal como herramienta de innovación docente. *RASE* vol.6; núm.1:114-21.
99. Peiró S, Portella E, El grupo nominal en el entorno sanitario. *Quaderns de Salut Publica i Administració de Serveis de Salut 1.* Valencia: Escola Valenciana d'Estudis per a la Salut; 2003.
100. Altman DG: *Practical statistics for medical research.* London: Chapman & Hall; 1999.

101. 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. *J Clin Epidemiol*. 2007;60(1):34-42.
102. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol*. 1993;46(12):1417-32.
103. 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 Health*. 2005;8:94-104.
104. Lujan-Tangarife JA, Cardona Arias JA. Construcción y validación de escalas de medición en salud: revisión de propiedades psicométricas. *Arch Med*. 2015;11(3):1.
105. Munro B. *Statistical methods for health care research*. Philadelphia (PA): JB Lippincott; 2000.
106. Michener LA, Leggin BG. A review of self-report scales for the assessment of functional limitation and disability of the shoulder. *J Hand Ther*. 2001;14(2):68-76.
107. 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*. 2012;21(4):651-7.
108. Cabero Almenara, J. y Llorente Cejudo, M. C. La aplicación del juicio de experto como técnica de evaluación de las tecnologías de la información (TIC). *Revista de Tecnología de Información y Comunicación en Educación*. 2013;7(2):11-22.

109. Malla F, Zabala I. La previsión del futuro en la empresa (III): el método Delphi. *Estudios Empresariales*. 1978;39:13-24.
110. Landeta J. El método Delphi: una técnica de previsión del futuro. Barcelona: Ariel; 2002.
111. Gómez-Valero S, García-Pérez F, Flórez-García MT, Miangolarra-Page JC. Assessment of cross-cultural adaptations of patient-reported shoulder outcome measures in Spanish: a systematic review. *Shoulder Elbow*. 2017;9(4):233-46.
112. Rosales RS, Delgado EB, Díez de la Lastra-Bosch I. Evaluation of the Spanish version of the DASH and carpal tunnel syndrome health-related quality-of-life instruments: cross-cultural adaptation process and reliability. *J Hand Surg Am*. 2002;27(2):334-43.
113. Hervás MT, Navarro Collado MJ, Peiró S, Rodrigo Pérez JL, López Matéu P, Martínez Tello I. Spanish version of the DASH questionnaire. Cross-cultural adaptation, reliability, validity and responsiveness. *Med Clin (Barc)*. 2006;127(12):441-7.
114. Mulero-Portela AL, Colón-Santaella CL, Cruz-Gómez C. Cross-cultural adaptation of the Disability of Arm, Shoulder, and Hand questionnaire: Spanish for Puerto Rico version. *Int J Rehabil Res*. 2009;32(4):287-93.
115. Cuesta-Vargas AI, Gabel PC. Cross-cultural adaptation, reliability and validity of the Spanish version of the Upper Limb Functional Index. *Health Qual Life Outcomes*. 2013;11:126.
116. Membrilla-Mesa MD, Tejero-Fernández V, Cuesta-Vargas AI, Arroyo-Morales M. Validation and reliability of a Spanish version of Simple Shoulder Test (SST-Sp). *Qual Life Res*. 2015;24(2):411-6.

117. Membrilla-Mesa MD, Cuesta-Vargas AI, Pozuelo-Calvo R, Tejero-Fernández V, Martín-Martín L, Arroyo-Morales M. Shoulder Pain and Disability Index: cross cultural validation and evaluation of psychometric properties of the Spanish version. *Health Qual Life Outcomes*. 2015;13(1):200.
118. Torres-Lacomba M, Sánchez-Sánchez B, Prieto-Gómez V, Pacheco-da-Costa S, Yuste-Sánchez MJ, Navarro-Brazález B, Gutiérrez-Ortega C. Spanish cultural adaptation and validation of the shoulder pain and disability index, and the oxford shoulder score after breast cancer surgery. *Health Qual Life Outcomes*. 2015;13:63.
119. Alvarez-Nemegyei J, Puerto-Ceballos I, Guzmán-Hau W, Bassol-Perea A, Nuño-Gutiérrez BL. Development of a Spanish-language version of the Shoulder Disability Questionnaire. *J Clin Rheumatol*. 2005;11(4):185-7.
120. Arcuri F, Barclay F, Nacul I. Traducción, adaptación trans-cultural, validación y medición de propiedades de la versión en español del índice Western Ontario Rotator Cuff (WORC). *Artroscopia*. 2015;22(2):56-60.
121. Arcuri F, Nacul I, Barclay F. Traducción, adaptación trans-cultural, validación y medición de propiedades de la versión en español del índice Western Ontario Shoulder Instability (WOSI). *Artroscopia*. 2015;22(4):125-33.
122. Arroyo Aljano R, González-Viejo MA. Validación al castellano del Wheelchair Users Shoulder Pain Index (WUSPI) *Rehabilitación*. 2009;43(1):2-9.
123. Arcuri F, Barclay F, Nacul I. Translation, cultural adaptation and validation of the Simple Shoulder Test to Spanish. *Orthop J Sports Med*. 2014, 12(2).
124. Espinosa-Cuervo G, Guillermin F, Rat AC, Duarte-Salazar C, Alemán-Hernández SI, Vergara-Álvarez Y, Goycochea-Robles MV. Transculturalization and validation of a Spanish translation of the specific lower limb osteoarthritis and

- quality of life questionnaire AMICAL: Arthrose des Membres Inférieurs et Qualité de vie AMIQUAL. *Reumatol Clin.* 2014;10(4):241-7.
125. Castellet E, Ares O, Celaya F, Valentí-Azcárate A, Salvador A, Torres A, Sesma P; SEROD group. Transcultural adaptation and validation of the "Hip and Knee" questionnaire into Spanish. *Health Qual Life Outcomes.* 2014;12:76.
126. Vaquero J, Longo UG, Forriol F, Martinelli N, Vethencourt R, Denaro V. Reliability, validity and responsiveness of the Spanish version of the Knee Injury and Osteoarthritis Outcome Score (KOOS) in patients with chondral lesion of the knee. *Knee Surg Sports Traumatol Arthrosc.* 2014;22(1):104-8.
127. Villanueva I, del Mar Guzman M, Javier Toyos F, Ariza-Ariza R, Navarro F. Relative efficiency and validity properties of a visual analogue vs a categorical scaled version of the Western Ontario and McMaster Universities Osteoarthritis (WOMAC) Index: Spanish versions. *Osteoarthritis Cartilage.* 2004;12(3):225-31.
128. Ares O, Castellet E, Maculé F, León V, Montañez E, Freire A, Hinarejos P, Montserrat F, Amillo JR. Translation and validation of 'The Knee Society Clinical Rating System' into Spanish. *Knee Surg Sports Traumatol Arthrosc.* 2013;21(11):2618-24.
129. González Sáenz de Tejada M, Escobar A, Herdman M, Herrera C, García L, Sarasqueta C. Adaptation and validation of the Osteoarthritis Knee and Hip Quality of Life (OAKHQOL) questionnaire for use in patients with osteoarthritis in Spain. *Clin Rheumatol.* 2011;30(12):1563-75.
130. Cruz-Díaz D, Lomas-Vega R, Osuna-Pérez MC, Hita-Contreras F, Fernández ÁD, Martínez-Amat A. The Spanish lower extremity functional scale: a reliable, valid and responsive questionnaire to assess musculoskeletal disorders in the lower extremity. *Disabil Rehabil.* 2014;36(23):2005-11.

131. Cuesta-Vargas AI, Gabel CP, Bennett P. Cross cultural adaptation and validation of a Spanish version of the Lower Limb Functional Index. *Health Qual Life Outcomes*. 2014;12:75.
132. Hernandez-Sanchez S, Hidalgo MD, Gomez A. Cross-cultural adaptation of VISA-P score for patellar tendinopathy in Spanish population. *J Orthop Sports Phys Ther*. 2011;41(8):581-91.
133. Gil-Gámez J, Pecos-Martín D, Kujala UM, Martínez-Merinerio P, Montañez-Aguilera FJ, Romero-Franco N, Gallego-Izquierdo T. Validation and cultural adaptation of "Kujala Score" in Spanish. *Knee Surg Sports Traumatol Arthrosc*. 2015.
134. Batlle-Guada E, Esteve-Vives J, Piera MC, Hargreaves R, Cuts J. Traducción y adaptación al español del cuestionario WOMAC específico para artrosis de rodilla y cadera. *Rev Esp Reumatol*. 1999;26:38-45.
135. Seijas R, Sallent A, Ruiz-Ibán MA, Ares O, Marín-Peña O, Cuéllar R, Muriel A. Validation of the Spanish version of the Hip Outcome Score: a multicenter study. *Health Qual Life Outcomes*. 2014 May 13;12:70.
136. Ruiz-Ibán MA, Seijas R, Sallent A, Ares O, Marín-Peña O, Muriel A, Cuéllar R. The international Hip Outcome Tool-33 (iHOT-33): multicenter validation and translation to Spanish. *Health Qual Life Outcomes*. 2015. 20;13:62.
137. Roddey TS, Cook KF, O'Malley KJ and Gartsman GM. The relationship among strength and mobility measures and self-report outcome scores in persons after rotator cuff repair surgery: impairment measures are not enough. *J Shoulder Elbow Surg*. 2005;14:95-8.

138. Greenhalgh J, Meadows K. The effectiveness of the use of patient-based measures of health in routine practice in improving the process and outcomes of patient care: a literature review. *J Eval Clin Pract.* 1999;5:401–16.
139. Wylie JD, Beckmann JT, Granger E and Tashjian RZ. Functional outcomes assessment in shoulder surgery. *World J Orthop.* 2014;5:623–33.
140. Angst F, Schwyzer HK, Aeschlimann A, Simmen BR and Goldhahn J. Measures of adult shoulder function: Disabilities of the Arm, Shoulder, and Hand Questionnaire (DASH) and its short version (QuickDASH), Shoulder Pain and Disability Index (SPADI), American Shoulder and Elbow Surgeons (ASES) Society standardized shoulder assessment form, Constant (Murley) Score (CS), Simple Shoulder Test (SST), Oxford Shoulder Score (OSS), Shoulder Disability Questionnaire (SDQ), and Western Ontario Shoulder Instability Index (WOSI). *Arthritis Care Res.* 2011;63:174-88.
141. Schmidt S, Ferrer M, González M, et al. Evaluation of shoulder-specific patient-reported outcome measures: a systematic and standardized comparison of available evidence. *J Shoulder Elbow Surg.* 2014;23:434-44.
142. St-Pierre C, Desmeules F, Dionne CE, FreÅLmont P, MacDermid JC and Roy JS. Psychometric properties of self-reported questionnaires for the evaluation of symptoms and functional limitations in individuals with rotator cuff disorders: a systematic review. *Disabil Rehabil.* 2016;38:103-22.
143. Huang H, Grant JA, Miller BS, Mirza FM and Gagnier JJ. A systematic review of the psychometric properties of patient-reported outcome instruments for use in patients with rotator cuff disease. *Am J Sports Med.* 2015;43:2572-82.
144. Collins NJ, Misra D, Felson DT, Crossley KM, Roos EM. Measures of knee function: International Knee Documentation Committee (IKDC) Subjective Knee

- Evaluation Form, Knee Injury and Osteoarthritis Outcome Score (KOOS), Knee Injury and Osteoarthritis Outcome Score Physical Function Short Form (KOOS-PS), Knee Outcome Survey Activities of Daily Living Scale (KOS-ADL), Lysholm Knee Scoring Scale, Oxford Knee Score (OKS), Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), Activity Rating Scale (ARS), and Tegner Activity Score (TAS). *Arthritis Care Res (Hoboken)*. 2011;11:208-28.
145. Rolfson O, Eresian Chenok K, Bohm E, L'ubbeke A, Denissen G, Dunn J, et al. Patient-reported outcome measures in arthroplasty registries. *Acta Orthop*. 2016;87:3-8.
146. White DK, Master H. Patient-reported measures of physical function in knee osteoarthritis. *Rheum Dis Clin North Am*. 2016;42:239-52.
147. Nilsson A, Bremander A. Measures of hip function and symptoms: Harris Hip Score (HHS), Hip Disability and Osteoarthritis Outcome Score (HOOS), Oxford Hip Score (OHS), Lequesne Index of Severity for Osteoarthritis of the Hip (LISOH), and American Academy of Orthopedic Surgeons (AAOS) Hip and Knee Questionnaire. *Arthritis Care Res (Hoboken)*. 2011;63:200-7.
148. Harris K, Dawson J, Gibbons E, Lim CR, Beard DJ, Fitzpatrick R, et al. Systematic review of measurement properties of patient-reported outcome measures used in patients undergoing hip and knee arthroplasty. *Patient Relat Outcome Meas*. 2016;7:101-8.
149. Ramisetty N, Kwon Y, Mohtadi N. Patient-reported outcome measures for hip preservation surgery-a systematic review of the literature. *J Hip Preserv Surg*. 2015;2:15-27.

150. Kemp JL, Collins NJ, Roos EM, Crossley KM. Psychometric properties of patient-reported outcome measures for hip arthroscopic surgery. *Am J Sports Med.* 2013;41:2065–73.
151. Thorborg K, Tjissen M, Habets B, Bartels EM, Roos EM, Kemp J, et al. Patient-Reported Outcome (PRO) questionnaires for young to middle-aged adults with hip and groin disability: a systematic review of the clinimetric evidence. *Br J Sports Med.* 2015;49:812.
152. Muñiz J, Elosua P, Hambleton RK. Directrices para la traducción y adaptación de los tests: segunda edición. *Psicothema.* 2013;25:151-7.
153. Muñiz J, Hambleton RK. Directrices para la traducción y adaptación de tests. *Papeles del Psicólogo.* 1996;66:63-70.
154. Bennett RE. Using new technology to improve assessment. *Educational Measurement: Issues and Practice.* 1999;18:5-12.
155. Breithaupt KJ, Mills CN, Melican GJ. Facing the opportunities of the future. In D. Bartram y R.K. Hambleton (Eds.), *Computerbased testing and the Internet.* West Sussex, UK: Wiley; 2006.
156. Drasgow F, Luecht RM, Bennett RE. Technology and testing. In RL Brennan (Ed.), *Educational measurement.* Westport, CT: ACE/Praeger; 2006.
157. Hambleton, R.K. Theory, methods, and practices in testing for the 21st century. *Psicothema.* 2004;16:696-701.
158. Byrne BM, Leong FT, Hambleton RK, Oakland T, van de Vijver FJ, Cheung FM. A critical analysis of cross-cultural research and testing practices: Implications for improved education and training in psychology. *Training and Education in Professional Psychology.* 2009;3:94-105.

159. Kennedy CA, Beaton DE, Smith P, et al. Measurement properties of the QuickDASH (disabilities of the arm, shoulder and hand) outcome measure and cross-cultural adaptations of the QuickDASH: a systematic review. *Qual Life Res* 2013;22:2509–47.
160. Varghese M, Lamb J, Rambani R and Venkateswaran B. The use of shoulder scoring systems and outcome measures in the UK. *Ann R Coll Surg Engl*. 2014;96:590–2.
161. Page MJ, McKenzie JE, Green SE, et al. Core domain and outcome measurement sets for shoulder pain trials are needed: systematic review of physical therapy trials. *J Clin Epidemiol*. 2015;68:1270–81.
162. Paul A, Lewis M, Shadforth MF, Croft PR, Van der Windt DA and Hay EM. A comparison of four shoulder-specific questionnaires in primary care. *Ann Rheum Dis*. 2004;63:1293–99.
163. Bellamy N. WOMAC Osteoarthritis Index user guide. London (Ontario, Canada): University of Western Ontario; 1995.
164. Bellamy N. WOMAC Osteoarthritis Index user guide. Version V. Brisbane (Australia): CONROD, The University of Queensland; 2002.
165. López Alonso SR, Martínez Sánchez CM, Romero Cañadillas AB, Navarro Casado F, González Rojo J. Propiedades métricas del cuestionario WOMAC y de una versión reducida para medir la sintomatología y la discapacidad física. *Aten Primaria*. 2009;41:613–20.
166. Escobar A, Vrotsoua K, Bilbao A, Quintana JM, García-Pérez L, Herrera-Espiñeira C. Validación de una escala reducida de capacidad funcional del cuestionario WOMAC. *Gac Sanit*. 2011;25:513–8.

167. Bilbao A, Quintana JM, Escobar A, Las Hayas C, Orive M. Validation of a proposed WOMAC short form for patients with hip osteoarthritis. *Health Qual Life Outcomes*. 2011;9:75.
168. Sánchez-Sotelo J. Instrumentos de valoración del estado de la salud en Traumatología y Cirugía Ortopédica. *Rev Ortop Traumatol*. 2004;48:304–14.
169. Gómez-Valero S, García-Pérez F, Flórez-García MT, Miangolarra-Page JC. Revisión sistemática de los cuestionarios autocumplimentados adaptados al español para la valoración funcional de pacientes con patología de cadera. *Reumatol Clin*. 2017;14:317-82.
170. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee Injury and Osteoarthritis Outcome Score (KOOS): development of a self-administered outcome measure. *J Orthop Sports Phys Ther*. 1998;28:88–96.
171. Thorborg K, Roos EM, Bartels EM, Petersen J, Hölmich P. Validity, reliability and responsiveness of patient-reported outcome questionnaires when assessing hip and groin disability: A systematic review. *Br J Sports Med*. 2010;44:1186–96.
172. Griffin DR, Parsons N, Mohtadi NG, Safran MR, Multicenter Arthroscopy of the Hip Outcomes Research Network. A short version of the International Hip Outcome Tool (iHOT-12) for use in routine clinical practice. *Arthroscopy*. 2012;28:611–6.
173. March Cerdá JC, Prieto Rodríguez MA, Hernán García M, Solas Gaspar O. Técnicas cualitativas para la investigación en salud pública y gestión de servicios de salud: algo más que otro tipo de técnicas. *Gac Sanit*. 1999;13:312-19.
174. Íñiguez Rueda L. Investigación y evaluación cualitativa: bases teóricas y conceptuales. *Aten Primaria*. 1999;23:496-502.

175. Jones J, Hunter D. Consensus methods for medical and health services research. *BMJ*. 1995;311:376-80.
176. Fleurence RL, Torgerson DJ. Setting priorities for research. *Health Policy*. 2004;69:1-10.
177. Moreno-Casbas T, Martín-Arribas C, Ortis-Cortes I, Comet-Cortes P; Investe-ISCIII Co-ordination and Development of Nursing Research Centre. Identification of priorities for nursing research in Spain: a Delphi study. *J Adv Nurs*. 2001;35:857-63.
178. Lionis C, Stoffers HE, Hummers-Pradier E, Griffiths F, Rotar-Palvic D, Rethans JJ. Settings priorities and identifying barriers for general practice research in Europe. Results from an EGPRW meeting. *Fam Pract*. 2004;21:587-93.
179. Ericsson KA, Charness N. Expert performance: its structure and acquisition. *Am Psychol*. 1994;49:725-47.
180. Norman G. Building on experience. The development of clinical reasoning. *N Engl J Med*. 2006;355:2251-2.
181. Kassirer JP. Teaching Clinical Reasoning: Case-Based and Coached. *Acad Med*. 2010;85:1118-24.
182. Bowen JL. Educational strategies to promote clinical diagnostic reasoning. *N Engl J Med*. 2006;355:2217-25.
183. Martínez Fernando, Caballero. Mejora de un método de consenso experto para reducir la variabilidad de la práctica clínica. Tesis doctoral. 2014. UAM. [file:///D:/caballero_martinez_fernando%20\(1\).pdf](file:///D:/caballero_martinez_fernando%20(1).pdf)

14. ANEXOS

14.1. Anexo 1. E-mail enviado para convocar a los participantes del grupo de expertos.

Buenos días,

Le envío este mail para invitarle a participar en una reunión de un grupo de expertos.

Hoy en día son muchas las escalas que nos encontramos en nuestro día a día y son tantas que muchas veces no sabemos cuáles y cómo usarlas.

Nuestro objetivo es daros a conocer los cuestionarios autoadministrados de valoración funcional del hombro (una de las articulaciones más demandadas en nuestra consulta) adaptados a la población española.

Y entre todos, analizar la calidad de los procedimientos de adaptación transcultural y sus propiedades clinimétricas, para así seleccionar la escala de hombro más adecuada para incorporarla al sistema informático del hospital, facilitando así su uso en la práctica clínica.

Para facilitar el trabajo, os adjuntamos como documentación la información sobre las escalas adaptadas al español para la valoración funcional de hombro adaptadas a la población española sobre las que se va a trabajar y las distintas escalas por separado, como apoyo, para valorar así el análisis de la calidad del proceso de adaptación y las propiedades de cada versión.

La reunión tendrá lugar el jueves 17 de marzo, en la sala de reuniones del servicio de Medicina Física y Rehabilitación, a las 14:00h.

Por favor, revise esta información para el día de la reunión y ruego de su consentimiento para la participación en dicha reunión.

Un saludo,

Sara Gómez Valero.

14.2. Anexo 2. Artículo de los cuestionarios autocumplimentados adaptados al español para la valoración funcional de pacientes con afecciones de hombro.



Assessment of cross-cultural adaptations of patient-reported shoulder outcome measures in Spanish: a systematic review

Shoulder & Elbow
2017, Vol. 9(4) 233–246
© The Author(s) 2017
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/1758573217694340
journals.sagepub.com/home/sel
SAGE

Sara Gómez-Valero¹, Fernando García-Pérez^{1,2}, Mariano Tomás Flórez-García^{1,2} and Juan Carlos Miangolarra-Page²

Abstract

Background: The present study aimed to conduct a systematic review of self-administered shoulder-disability functional assessment questionnaires adapted to Spanish, analyzing the quality of the transcultural adaptation and the clinimetric properties of the new version.

Methods: A search of the main biomedical databases was conducted to locate Spanish shoulder function assessment scales. The authors reviewed the papers and considered whether the process of adaptation of the questionnaire had followed international recommendations, and whether its psychometric properties had been appropriately assessed.

Results: The search identified nine shoulder function assessment scales adapted to Spanish: Disabilities of the Arm, Shoulder and Hand Questionnaire (DASH), Upper Limb Functional Index (ULFI), Simple Shoulder Test (SST), Shoulder Pain and Disability Index (SPADI), Oxford Shoulder Score (OSS), Shoulder Disability Questionnaire (SDQ), Western Ontario Rotator Cuff index (WORC), Western Ontario Shoulder Instability index (WOSI) and Wheelchair Users Shoulder Pain Index (WUSPI). The DASH was adapted on three occasions and the SPADI on two. The transcultural adaptation procedure was generally satisfactory, albeit somewhat less rigorous for the SDQ and WUSPI. Reliability was analyzed in all cases. Validity was not measured for one of the adaptations of the DASH, nor was it measured for the SDQ.

Conclusions: The transcultural adaptation was satisfactory and the psychometric properties analyzed were similar to both the original version and other versions adapted to other languages.

Keywords

cross-cultural adaptation, outcome assessment, questionnaire, shoulder, Spanish version

Date received: 20th September 2016; revised: 28th November 2016; accepted: 13th January 2017

Introduction

At any given time, shoulder pain may affect up to 18% to 26% of adult individuals,¹ with the most usual course being rotator cuff disorders.² The prevalence of shoulder problems was 17.12% in a primary health-care setting in Spain.³

More than 40 instruments have been developed for measuring treatment outcomes in shoulder disorders.^{4–6} Data drawn from physical examinations and/or complementary tests must be completed with assessment of the impact of the disorder from the patient's point of view (pain, ability to perform certain activities, etc.). The subjective impact of any given disease can vary

from one patient to another. Hence, a specific shoulder disorder may affect functional capacity, work performance and/or quality of life differently according to the patient concerned. For example, among older

¹Department of Physical Medicine and Rehabilitation, Hospital Universitario Fundación Alcorcón, Alcorcón, Spain

²Department of Physical Therapy, Occupational Therapy, Physical Medicine and Rehabilitation, Universidad Rey Juan Carlos, Alcorcón, Spain

Corresponding author:

Sara Gómez-Valero, Hospital Universitario Fundación Alcorcón, c/ Budapest, Alcorcón, 28922 Spain.
Email: saragova@gmail.com

individuals, rotator cuff tears may or may not impact their ability to perform daily functional tasks, whereas, in young people, glenohumeral instability can sometimes be a disease that restricts upper limb activity. Such data, namely patient-reported outcomes, can be obtained by means of self-administered questionnaires. There are over 30 questionnaires,⁷ none of which has shown itself to be superior to the others.⁸ Some are applicable to any upper limb disorder and assess it as a whole, regarding it as a single functional unit. There are other questionnaires exclusively designed for the shoulder regions, whether general (for any disease or disorder) or specific (for certain disorders). There are also questionnaires designed for certain subgroups of patients, such as sportspersons and wheelchair users.

It is important to choose questionnaires that are valid, reliable and sensitive to clinical changes. Most have been developed in English-speaking countries for Anglosphere cultures. Hence, if a questionnaire has not been drawn up in the country, geographical region, culture or language in which it is to be applied, an appropriate transcultural adaptation must be made to ensure that it retains the same meaning as the original,^{9,10} and so prevents erroneous interpretations as a result of cultural or life-style differences. This means, that transcultural adaptation is a more complex process than simple direct or literal translation and, in addition, is no guarantee of the preservation of the psychometric properties in the adapted questionnaire. Accordingly, the psychometric properties of the new version must always be analyzed to ensure that it is a tool equivalent to the original, without any assumptions being made in this respect. In daily practice, healthcare professionals need to assess their results and compare the effectiveness of a given treatment against that reported by other studies. Culturally adapting an existing questionnaire, in preference to developing a new one, is not only more economical, but also may facilitate future comparisons among different populations, provided that conceptual equivalence is successfully achieved. Transcultural adaptation is complex process because it requires changes to both the wording and structure of some questions for the questionnaire to be applied, on an equivalent basis, in populations other than those for which it was created. The use of questionnaires that are not equivalent to the original may give rise to unreliable to confusing results, which could limit the exchange of information among the scientific community.

Spanish is the second most widely spoken language in the world and is witnessing a rise in the number of speakers. In 2015, close on 470 million persons spoke it as their native language.¹¹ Not only is it spoken in Spain and Latin America, with small variations, but also it is habitually used in many other countries. For

example, in the USA, over thirty-six million people speak it as their mother tongue. It is thus necessary to have questionnaires adapted to Spanish to ensure that their use is equivalent to that of the original. A systematic review of self-administered shoulder questionnaires adapted to Portuguese was published recently.¹² It was found that, although most of the shoulder disability questionnaires had been properly adapted into Brazilian-Portuguese, some of them had either been inadequately tested or had not been tested at all. In 2013,¹³ questionnaires adapted to Spanish for patients with cervical and lumbar pain were reviewed: no similar reviews of shoulder disorder questionnaires have been published in Spanish.

The present study aimed to conduct a systematic review, by making an exhaustive bibliographic search of self-administered shoulder questionnaires adapted to Spanish, and analyzing both the methodological quality of the transcultural adaptation process and the psychometric properties of the Spanish-language versions.

Methods

Search strategy

We carried out a systematic review of papers that addressed both the transcultural adaptation to Spanish of self-administered questionnaires designed to assess patients with shoulder disabilities, and the process used to validate the resulting Spanish versions. A bibliographic search was made of international electronic databases (MEDLINE, Excerpta Medica dataBASE/EMBASE, Cumulative Index to Nursing and Allied Health Literature/CINAHL and Web of Science) from their date of creation until 31 December 2015. In addition, we searched other databases and directories with bibliographies in Spanish, namely, BIREME (SciELO, Lilacs and IBECs), *MEDICINA en Español* (MEDES), Allied and Complementary Database (AMED), IME and DIALNET. The following search terms and strategy were used in MEDLINE: 'Outcome' or 'Questionnaire' or 'Assessment' or 'Instruments' and 'Shoulder' or 'Rotator cuff' or 'Upper limb' and 'Spanish' or 'Spanish version' or 'Spanish validation' or 'Spanish translation' or 'Cross-cultural adaptation' or 'Cross-cultural validation'. Similar terms were used to search in other databases. Where necessary, search terms were modified with the assistance of an experienced medical librarian to reflect database vocabularies. In addition, we conducted a manual search using the names of the different shoulder scales as the key word; conducted a search on the Internet, including Google Scholar, to cover other types of publications; and manually examined the references in the papers located.

Selection criteria

Two of the authors reviewed the title and abstract of every paper retrieved. When this suggested that a given paper might be eligible, they then read the full text. We included papers, without language restrictions, reporting observational studies that described the process of transcultural adaptation to Spanish of self-administered questionnaires designed to carry out a functional assessment of patients with shoulder disabilities, and that also analyzed the psychometric properties of the new version. Papers that only analyzed the properties of a previously adapted questionnaire, research protocols and abstracts of conferences were excluded.

Data analysis

Based on the papers selected, two authors structured and analyzed the results of the manuscript. Where any differences of opinion could not be settled, the opinion of a third reviewer was sought. The analysis of the information is described below.

Characteristics of study participants. Data were collected on patients who participated in the study, including country and town, total number of patients included, diagnosis, age, sex, and time taken to complete the questionnaire. Studies were checked to see whether they covered at least 50 patients, which is the ideal minimum number advisable for transcultural adaptation studies.^{14,15}

Assessment of the methodology used to carry out the transcultural adaptation. A check was carried out to ensure that the methodology coincided with the five steps usually recommended in the international bibliography,^{9,16-20}. These steps are: (1) direct translation of the original questionnaire to Spanish (by at least two, independent, bilingual translators whose native language is Spanish); (2) synthesis of translations and resolution of possible disagreements; (3) inverse or back translation (of the consensus Spanish translation into the original language by at least two independent translators who were blinded to the original questionnaire); (4) review by an expert committee to consolidate and develop the pre-final version, ensuring the semantic, idiomatic, experiential and conceptual equivalence of the scale; and (5) pilot testing of the pre-final questionnaire with Spanish-speaking subjects (ideally 30-40), searching for unanswered items and possible problems of comprehension). At each step, we considered whether or not the procedure had been correctly performed (or whether the correction was doubtful), whether it had not been performed, or whether the authors had failed to provide the necessary information.

Evaluation of the psychometric properties of the Spanish-language version. The aspects analyzed were: reliability, validity and sensitivity to change.^{9,15,20-23} Reliability includes internal consistency, test-retest reproducibility, and agreement. Internal consistency is calculated by means of Cronbach's alpha (α). Where the instrument is made up of subscales, there is the need to calculate the correlation between the items making up each domain and the total scale. Factor analysis is applied to determine the dimensionality of the items. Test-retest reproducibility is calculated using the intraclass correlation coefficient (ICC). Agreement is assessed using the standard error of measurement (SEm) and minimal detectable change (MDC₉₀). Validity is measured using correlation coefficients, such as Pearson's or Spearman's, and correlation can be direct, indirect or absent. Floor and ceiling effects refer to the percentage of subjects who obtain the lowest and highest possible scores. This should not exceed 15% because, otherwise, the content validity of the scale will be limited because patients with extreme values cannot be distinguished from one another. Sensitivity is the ability of an instrument to detect important clinical changes across time. The minimum detectable change is mainly measured by reference to effect size (ES) and the standardized response mean (SRM).

Assessment of the direct applicability of scales adapted to one Spanish-speaking country in terms of their possible use in another. We analyzed whether adaptations made in Latin American countries needed some modification to be used in Spain, and vice versa. Despite the fact that the great majority of Spanish words are universal, to ensure comprehension, we assessed whether certain terms needed changing, in cases where they might be seldom used or undergo shifts in nuance or meaning from one culture to another.

Results

The bibliographic search identified eleven relevant papers,²⁴⁻³⁴ Eight²⁴⁻³¹ of these were identified via MEDLINE and the remaining three³²⁻³⁴ were via other sources (two on the Internet and one in the CINAHL database). The manual search failed to locate any relevant papers. These eleven papers yielded nine self-administered questionnaires that had been translated and transculturally adapted to the Spanish population, and in respect of which the clinimetric properties of the new version had been studied. One paper³⁰ published the adaptation of two different questionnaires, thereby giving a total of twelve studies in the eleven papers. In all cases, the original questionnaire had been drawn up in English. There were two

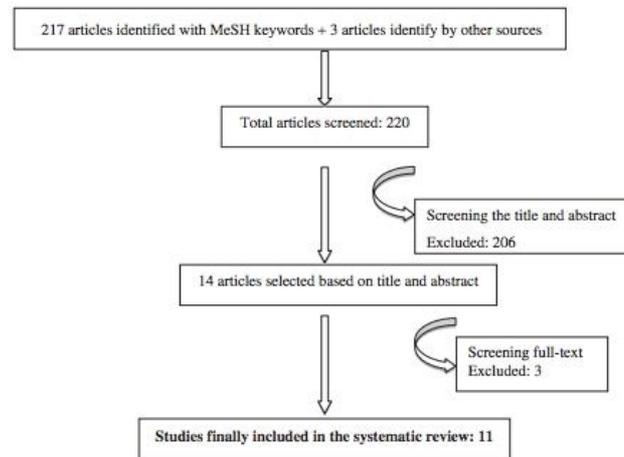


Figure 1. Flow diagram demonstrating the selection of studies for inclusion. MeSH: Medical Subject Headings.

questionnaires applicable to any type of upper limb disorder: the Disabilities of the Arm, Shoulder and Hand Questionnaire (DASH)²⁴⁻²⁶ and Upper Limb Functional Index (ULFI).²⁷ There were four general shoulder questionnaires: the Simple Shoulder Test (SST),²⁸ Shoulder Pain and Disability Index (SPADI),^{29,30} Oxford Shoulder Score (OSS)³⁰ and Shoulder Disability Questionnaire (SDQ).³¹ Finally, there were three questionnaires for specific situations: the Western Ontario Rotator Cuff Index (WORC)³² for patients with rotator cuff disease, the Western Ontario Shoulder Instability Index (WOSI)³³ for patients with shoulder instability and the Wheelchair Users Shoulder Pain Index (WUSPI)³⁴ for patients confined to wheelchairs. The DASH had been adapted and validated on three occasions: two in Spain^{24,25} and one in Puerto Rico.²⁶ The SPADI had been adapted and validated on two occasions: both in Spain^{29,30} and practically simultaneously; one of these two adaptations had been carried out jointly with the OSS questionnaire, and both had been published in a single paper.³⁰ In all, there were twelve adaptations for nine questionnaires. Our search located no questionnaire that had been originally developed in Spanish.

Table 1 shows the demographic and clinical characteristics of the respective populations targeted by the twelve studies. The papers included the adapted scale and instructions for patients in all cases, save for three questionnaires: the adaptation of the DASH in Puerto Rico,²⁶ the OSS³⁰ and the WORC.³² Most of the papers had been published very recently. Six had been published in 2015^{28-30,32,33} and one in 2013.²⁷ Four scales

were adapted in Latin American countries^{26,31-33} and the remainder in Spain.^{24,25,27-30,34} In eight of the twelve adaptations,^{25,27-30,32,33} the number of patients exceeded the ideal minimum of 50 recommended for these types of studies. Only one of the three adaptations of the DASH scale, namely the version produced by Hervás et al.,²⁵ included patients with shoulder disabilities; the other two were validated in patients with carpal tunnel syndrome,²⁴ and the version for Puerto Rico was validated in patients after breast cancer surgery.²⁶ The authors who simultaneously adapted the SPADI and OSS questionnaires only included patients intervened for breast cancer;³⁰ in the other adaptation of the SPADI,²⁹ patients with shoulder disorders were included. The adaptation of the WUSPI exclusively included spinal cord injury patients who were wheelchair users.³⁴ The time taken to complete the questionnaire was recorded in only four cases^{25,30,34} and proved to be similar to that of the original version.

The DASH has two additional optional modules, each comprising four items, which were adapted in the version by Hervás et al.²⁵ and in that produced in 2009 for Puerto Rican patients.²⁶ Each of the original questionnaires specified the period of time preceding the questionnaire completion date that patients were to take into account when answering. However, in the adaptation of the WUSPI³⁴ and in one of the adaptations of the SPADI,³⁰ the adapted scale failed to give any clear indication of the period of time that patients were expected to analyze when assessing their shoulder status. In general, the transcultural adaptation process of all scales required modifications, albeit of little

Table 1. Demographic and clinical characteristics of the translocatal-adaptation study populations.

Scale	Author (year of adaptation)	Country (city or town)	Total number of patients included	Diagnosis	Age (years)	Percentage of women	Completion time (minutes)
DASH ²⁴	Rosales et al. (2002)	Spain (Tenerife)	42	Carpal tunnel syndrome	Mean: 54 Range: 34–63	85.7%	Not measured
DASH ²⁵	Herrás et al. (2006)	Spain (Valencia)	98	Different upper limb disorders (73.5% with shoulder problems, essentially: 34.6% subacromial impingement syndrome; 27.3% tendinitis; and 9.1% humerus or scapular (shoulder blade) fractures)	Range: 27–76 (over 50% 25–65)	64.3 %	10
DASH ²⁶	Mulero-Portela et al. (2009)	Puerto Rico (San Juan)	44	After breast cancer surgery (patients who had undergone operations in the previous 3 years and completed their treatment)	Mean: 52.59 Range: 34–84	100%	Not measured
ULFI ²⁷	Cuesta-Vargas y Gabel (2013)	Spain (Málaga)	126	Different upper limb disorders (56% with shoulder problems, i.e. subacromial impingement syndrome or tendinitis)	Mean: 49 (SD 21)	54.8 %	Not measured
SST ²⁸	Membrilla-Mesa et al. (2015)	Spain (Granada)	66	Shoulder pain as a result of different disorders, particularly: 31% humerus fractures; 21.23% calcific tendinitis; and 19.74% rotator cuff tear	Mean: 51 (SD 12)	59.13 %	Not measured
SPADI ²⁹	Membrilla-Mesa et al. (2015)	Spain (Málaga)	136	Shoulder pain as a result of different disorders, particularly: 22.1% bicipital tendinitis, 21.3% humerus fractures, 17.6% rotator cuff tear and 11.8% calcific tendinitis)	Mean: 49.8 (SD 15)	55.1%	Not measured
SPADI ³⁰	Torres-Lacomba et al. (2015)	Spain (Alcalá de Henares)	120	After breast cancer surgery (patients with clinical course of less than 6 months)	Mean: 54.2 (SD 11)	100 %	3 (SD 1.9)

(continued)

Table 1. Continued

Scale	Author (year of adaptation)	Country (city or town)	Total number of patients included	Diagnosis	Age (years)	Percentage of women	Completion time (minutes)
OSS ⁶	Torres-Lacomba et al (2015)	Spain (Alcalá de Henares)	120	After breast cancer surgery (patients with clinical course of less than 6 months)	Mean: 54.2 (SD 11)	100 %	3.4 (SD 1.4)
SDQ ²¹	Álvarez-Nemegyei et al (2005)	Mexico (Yucatán)	35	Shoulder pain as a result of subacromial impingement syndrome	Mean: 55 (SD 9)	77.1%	Not measured
WORC ²²	Arcuri et al (2015)	Argentina (Buenos Aires)	60	Tendinitis or rotator cuff tear	Mean: 57 (SD 12.3) Range: 19–76	44 %	Not measured
WOSI ²³	Arcuri et al (2015)	Argentina (Buenos Aires)	60	Shoulder instability	Mean: 40.12 (SD 17) Range: 17–65	45%	Not measured
WUJSP ²⁴	Amayo-Aljaro y González-Veijo (2009)	Spain (Barcelona)	42	Spinal cord injuries (28.9% cervical, 36.8% upper dorsal and 34.2% lower dorsal)	Median: 40 Range: 15–69	19 %	5

DASH, Disabilities of the Arm, Shoulder and Hand Questionnaire; UIFI, Upper Limb Functional Index; SST, Simple Shoulder Test; SPADI, Shoulder Pain and Disability Index; OSS, Oxford Shoulder Score; SDQ, Shoulder Disability Questionnaire; WORC, Western Ontario Rotator Cuff Index; WOSI, Western Ontario Shoulder Instability Index; WUJSP, Wheelchair Users Shoulder Pain Index.

Table 2. Assessment of the methodology used for transcultural adaptation of questionnaires.

Scale	Translation	Summary	Back translation	Analysis by expert committee	Pilot test (number of patients)
DASH ²⁴ Rosales et al. (2002)	+	+	+	+	+ (50)
DASH ²⁵ Hervás et al. (2006)	+	+	+	+	+ (15)
DASH ²⁶ Muleto-Portela et al. (2009)	+	+	+	+	+ (32)
ULFI ²⁷	+	+	+	?	-
SST ²⁸	+	+	+	?	-
SPADI ²⁹ Membrilla-Mesa et al. (2015)	+	+	+	+	+ (40)
SPADI ³⁰ Torres-Lacomba et al. (2015)	+	+	+	+	+ (20)
OSS ³⁰	+	+	+	+	+ (20)
SDQ ³¹	-	0	-	0	-
WORC ³²	+	+	+	+	+ (10)
WOSI ³³	+	+	+	+	+ (10)
WUSPI ³⁴	?	0	?	0	-

DASH, Disabilities of the Arm, Shoulder and Hand Questionnaire; ULFI, Upper Limb Functional Index; SST, Simple Shoulder Test; SPADI, Shoulder Pain and Disability Index; OSS, Oxford Shoulder Score; SDQ, Shoulder Disability Questionnaire; WORC, Western Ontario Rotator Cuff Index; WOSI, Western Ontario Shoulder Instability Index; WUSPI, Wheelchair Users Shoulder Pain Index; (+), Correctly done; (?), Doubtful; (-), Incorrectly done or not done; (0), No information given as to whether or not done.

importance, with respect to the original version. In the three adaptations of the DASH,²⁴⁻²⁶ no identical questionnaires were obtained. In the version for Puerto Rico,²⁶ some Anglicisms had to be included in view of the fact that the country's second language is English. Similarly, the two adaptations of the SPADI^{29,30} were not identical. Even so, the different variants in the two questionnaires were equivalent.

The main results of the evaluation of the methodology used for transcultural adaptation are analyzed in Table 2 (adapted from Puga et al.¹¹). The three versions of the DASH,²⁴⁻²⁶ the two adaptations of the SPADI,^{29,30} and the OSS,³⁰ WORC³² and WOSI³³ versions rigorously complied with the international recommendations. The SDQ³¹ and WUSPI³⁴ were the least strict in this respect. Although pilot testing was performed for eight scales,^{24-26,29,30,32,33} only three^{24,26,29} included more than 30 subjects. The questionnaires were well accepted by patients, who reported no difficulties in understanding items or completing them.

The main psychometric properties analyzed in the Spanish versions are listed in Table 3. No study assessed all the metric properties of the new version. In general, the conclusion of the authors of each of the adaptations was that the psychometric properties assessed were acceptable and comparable to those of the original versions.

In terms of reliability, internal consistency was assessed in all scales, test-retest reproducibility in all but one²⁶ and agreement in only five.^{27-29,32,33} When the test-retest was performed, there were only six scales,^{24,25,27,29,31,33} which exceeded the ideal number of patients recommended for this assessment (i.e. 29). Total internal scale consistency was excellent for seven questionnaires,^{24-27,30-33} good for the WUSPI³⁴ and acceptable for the SST.²⁸ In the case of the adaptation of the SPADI by Membrilla-Mesa et al.,²⁹ it was good for one of the factors and excellent for the other. The test-retest time used was too long for the WORC³² and WOSI questionnaires³³ (up to 21 days), and too short

Table 3. Principal psychometric properties analyzed in the adapted questionnaires.

Scale	Total internal consistency (Cronbach's α)	Test-retest reproducibility: number of participants (time between two evaluations) ICC	Instrument of comparison: validity (correlation coefficient) (negative values indicate inverse correlation)	Floor/ceiling effects	Sensitivity ES SRM
DASH ¹⁴ Rosales et al. (2002)	0.95	42 (7 days) 0.97 (95% CI = -1.98 to 0.52)	Not used (not calculated)	-	-
DASH ¹⁵ Hervás et al. (2006)	0.96	30 (7 days to 10 days) 0.96, $p < 0.01$	SF-36: physical functioning subscale: -0.509 and pain subscale: -0.693. Significant correlation ($p < 0.01$) with all SF-36 dimensions in the direction expected under the initial hypothesis	-	1.03 0.86
DASH ¹⁶ Mulero-Portada et al. (2009)	0.97 (item range: 0.44 to 0.85)	Not analyzed	FACTB: (range 0.0% to 0.682). Significant correlation ($p < 0.01$) with four of the five FACTB subscales	No/No	-
ULFI ¹⁷	0.94 (item range: 0.92 to 0.96)	35 (7 days) 0.93 (range 0.92 to 0.95)	EQ-5D-3L: -0.59 (for the pain subscale: -0.51) QuickDASH: not valid (more than 30% of questionnaires with excessive missing answers) Factor structure analysis: unidimensional	-	-
SST ¹⁸	0.793 (item range: 0.667 to 0.723)	21 (2 days) 0.912 (range 0.687 to 0.944)	DASH: -0.73 VAS pain: -0.53 SF-12: physical functioning subscale -0.47 and mental subscale -0.43, con $p < 0.001$ Factor structure analysis: tridimensional	No/No	-

(continued)

Table 3. Continued

Scale	Total internal consistency (Cronbach's α)	Test-retest reproducibility: number of participants (time between two evaluations) ICC	Instrument of comparison: validity (correlation coefficient) (negative values indicate inverse correlation)	Floor/celling effects	Sensitivity ES SRM
SPADI ²⁹ Membriñilla-Mesa et al. (2015)	0.92 (95% CI = 0.91 to 0.95) and 0.82 (95% CI = 0.76 to 0.86) for each of the two factors	56 (2 days) 0.91 (95% CI = 0.88 to 0.94) with item range of 0.89 to 0.93	Significant correlations ($p < 0.01$) in the direction expected under the initial hypothesis DASH: pain subscale 0.80 and disability subscale 0.76 SF-12: total 0.40 (with the physical functioning and mental subscales) SST: pain subscale -0.71 and disability subscale -0.75 VAS pain: pain subscale 0.67 and disability subscale 0.65 Factor structure analysis: bidimensional	-	-
SPADI ³⁰ Torres-Lacomba et al. (2015)	0.965 (0.931 on the pain subscale and 0.953 on the function subscale)	20 (2 days) 0.992 (0.986 in the pain subscale and 0.991 on the function subscale) with $p < 0.01$	OSS: -0.674 FACT-B: -0.298 SDQ: 0.432 SF-36: -0.452 to -0.315 (for the physical functioning, role-physical, bodily pain and role-emotional dimensions) with $p < 0.01$	No/No	Global scale 0.59 0.82 Pain subscale 0.59 0.78 Function subscale 0.82 1.13 $p < 0.0001$
OSS ¹⁰	0.947	20 (2 days) 0.974	SPADI: -0.674 (pain subscale -0.640 and disability subscale -0.645) FACT-B: -0.343 SDQ: -0.469 SF-36: 0.312 to 0.391 (for the physical functioning, role-physical, bodily pain and role-emotional dimensions) with $p < 0.01$	No/No	-0.50 -0.70 $p < 0.0001$
SDQ ³¹	0.99	35 (1 hour) 0.99 (95% CI = 0.998 to 0.999)	Not used (not calculated)	-	-

(continued)

Table 3. Continued

Scale	Total internal consistency (Cronbach's α)	Test-retest reproducibility: number of participants (time between two evaluations) ICC	Instrument of comparison: validity (correlation coefficient) (negative values indicate inverse correlation)	Floor/ceiling effects	Sensitivity ES SRM
WORC ³²	0.952 (range among the 5 domains: 0.824 to 0.953)	10 (14 days to 21 days, median 14) 0.909 (95% CI = 0.841 to 0.949)	SST: 0.756 Constant Scale: 0.60	No/No	
WOSI ³³	0.948 (range among the 4 domains: 0.783 to 0.906)	60 (14 days to 21 days, median 14) 0.851 (95% CI = 0.587 to 0.879)	SST: 0.498 Constant Scale: 0.60	No/No	
WUSPI ³⁴	0.88	8 (not specified) 0.96	VAS pain (in 8 of 45 patients): 0.90	Yes/No	

DASH, Disabilities of the Arm, Shoulder and Hand Questionnaire; ULB, Upper Limb Functional Index; SST, Simple Shoulder Test; SPADI, Shoulder Pain and Disability Index; OSS, Oxford Shoulder Score; SDQ, Shoulder Disability Questionnaire; WORC, Western Ontario Rotator Cuff Index; WOSI, Western Ontario Shoulder Instability Index; WUEPI, Wheelchair Users Shoulder Pain Index; ICC, Intraclass correlation coefficient; CI, confidence interval; SF-36, Short Form 36 Health Survey; FACT-B, Functional Assessment of Cancer Therapy-Breast EQ-SD-3L, European Health Questionnaire 5 Dimensions 3 Levels; VAS, Visual analog scale; SF-12, Short Form 12 Health Survey; ES, effect size; SRM, standardized response mean.

for the SDQ³¹ (1 hour). The ICC was very good in all cases. Agreement was analyzed in five questionnaires.^{27-29,32,33} In the adaptation to Spanish of the SPADI, Membrilla-Mesa et al.²⁹ only analyzed the MDC₉₅ value, which was 12.2%. In the other four questionnaires, both the SEM and MDC₉₀ values were analyzed, with the respective results for the two being: ULFI, 3.52% and 8.03% (equivalent to 2 points of the maximum scale score of 25);²⁷ SST, 2.21% and 6.2%;²⁸ WORC 192.76 (domain range 39.43–63.82) and 533 (domain range 109.21–176.78) for a total scale score range of 0–2100;³³ and WOSI, 192.76 (domain range 109.21–176.78) and 533 (domain range 109.21–176.78) for a total scale score range of 0–2100.³³

Validity was assessed in all questionnaires, with the exception of the adaptations of the DASH in Puerto Rico²⁶ and the SDQ in Mexico.³¹ Validity varied according to the comparison instruments used but was generally adequate. The comparators used to validate the WORC and WOSI in Argentina^{32,33} were the Constant Scale and a version of the SST adapted to the Argentine population, for which there are no data published in the form of a paper.³⁵ Construct validity was only properly assessed in the adaptations of the DASH by Hervás et al.²⁵ and the SPADI by Membrilla-Mesa et al.,²⁹ in that there was a prior hypothesis regarding the correlation of the instruments being compared. Factor structure analysis was performed for only three questionnaires, the ULFI,²⁷ the SST²⁸ and one of the two adaptations of the SPADI,²⁹ with all being adapted by the same work group and the scales proving unidimensional in the first case, tridimensional in the second case and bidimensional in the third case. Only in two cases^{27,29} was the number of patients used appropriate (> 100). Floor and ceiling effects were not analyzed in five questionnaires,^{24,25,28,29,31} and, in the remainder, the WUSPI alone displayed evidence of a floor effect.³⁴

Sensitivity was examined in only three scales.^{25,30} ES and SRM values were obtained for the DASH adapted by Hervás et al.,²⁵ indicating high sensitivity in both cases. For one of the adaptations of the SPADI,³⁰ these values were moderate–high, respectively, for the global scale, and high in both cases for the functional subscale. The ES and SRM values for the OSS questionnaire³⁰ were low.

The clinimetric properties of the two optional DASH modules could not be analyzed in the adaptation of Hervás et al.,²⁵ as a result of the low number of respondents, and were similarly not analyzed in the version by Mulero-Portela et al.²⁶ for Puerto Rican patients.

In four questionnaires,^{26,31-33} the transcultural adaptations were produced in Latin American countries. In three of these cases,³¹⁻³³ none of the cultural

flaws were so marked as to prevent the questionnaire being used in Spain. The introduction of small changes in some of the vocabulary would permit the instruments to be used without problem in Spain: for example, in the adaptation of the SDQ for Mexico³¹ 'auto' could be replaced by 'coche' and 'manubrio' by 'manillar'. Similarly, in the case of the adaptations produced in Spain, which might contain words that were inappropriate for subjects of Latin American origin, these could be modified (e.g. by replacing 'coger' with 'agarrar') and, equally, be applicable there. In the case of the adaptation of the DASH to the Puerto Rican population,²⁶ it was necessary to permit some Anglicisms, which would render it inapplicable to Spain. In Spain, either of the other two versions of the DASH could be used.^{24,25}

Discussion

Outcome measures developed from the patient's standpoint are important for assessing the effectiveness of treatments administered. Indeed, clinician-made measurements correlate poorly with disability as perceived by patients, which, moreover, is the aspect of most relevance to them.³⁶ At an international level, numerous questionnaires have been developed for assessing patients' self-perceived impact of their shoulder disorders. There are more than 30 such questionnaires, and they are being increasingly used: each has advantages and disadvantages with respect to the rest, and none is superior to any other. Interest in this issue is growing and a number of reviews have been published in recent years.^{6-8,22,37-47} Most of the questionnaires have been developed in English-speaking countries. Spanish is the second most widely spoken language in the world, with widespread geographical coverage,¹¹ encompassing countries that use it as an official language, as well as cases where it is spoken by immigrants who reside in countries having other mother tongues. Adapting an already existing scale and ensuring that the new version maintains the psychometric properties of the original, is preferable to creating a wholly new scale. The latter option entails a greater effort in terms of cost and time, and increases the diversity of questionnaires. A single scale facilitates comparisons across different populations. Transcultural adaptation must ensure conceptual equivalence with the original version.

The goal of the present study was to conduct a systematic review of self-administered shoulder questionnaires adapted to Spanish, analyzing both the methodological quality of the transcultural adaptation process and the psychometric properties of the new version obtained. We located eleven papers²⁴⁻³⁴ on nine questionnaires, six of which had been published in 2015.^{28-30,32,33} Two were applicable to the entire upper limb: the DASH (adapted on three

occasions)²⁴⁻²⁶ and the ULFI.²⁷ Four were general shoulder questionnaires: the SST,²⁸ SPADI (adapted on two occasions),^{29,30} OSS³⁰ and SDQ.³¹ Lastly, there were three questionnaires for specific situations: the WORC³² for patients with rotator cuff disease, the WOSI³³ for cases of instability and the WUSPI³⁴ for wheelchair users. Of these, the most widely used at an international level are the DASH, SST, SPADI and OSS.

In the adaptation of three questionnaires only women (intervened for breast cancer) were included in two,^{26,30} and patients with carpal tunnel syndrome in another.²⁴ This had no relevant influence on the quality of the transcultural adaptation or on the psychometric properties of the questionnaires adapted. Similarly, the fact that four studies^{24,26,31,34} had fewer than 50 participants, a number lower than the ideally recommended minimum for transcultural adaptation studies, had no special significance.¹⁷

There is still no clear international consensus on the optimal approach to carrying out transcultural adaptation:¹⁰ indeed, there are over 30 methods and no 'gold standard'. What does appear to be agreed is that inverse (or back) translation and pilot testing are essential.¹⁰ The internationally recognized criteria for the process of adaptation to Spanish were followed with maximum rigor in the three versions of the DASH,²⁴⁻²⁶ the two adaptations of the SPADI,^{29,30} and the OSS³⁰ and WOSI³¹ questionnaires. By contrast, there were important shortcomings in the adaptations of the SDQ³¹ and WUSPI.³⁴ In general, there were no unduly pronounced cultural flaws, except in the case of the Puerto Rican population in the adaptation of the DASH,²⁶ where some Anglicisms had to be allowed. In the other three questionnaires adapted in Latin American countries,³¹⁻³³ the introduction of small changes in certain words used would make it possible for these to be used without any problem of comprehension in Spain. Conversely, the same could be carried out with questionnaires adapted in Spain.

When it comes to choosing which questionnaire to use in a specific situation, consideration must be given to methodological criteria (psychometric properties) and practices (time needed for completion and scoring; usefulness for specific disorders). No study assessed all the possible psychometric properties. Reliability was analyzed in a number of ways in all the questionnaires, and validity in all but one of the adaptations of the DASH²⁴ and SDQ.³¹ In the case of reliability, internal consistency was assessed in all scales, test-retest reproducibility in all but one²⁶ and agreement in only five.^{27-29,32,33} The psychometric properties analyzed were acceptable and generally similar to those of the original version and other scale versions adapted to other languages.

The DASH is the most widespread and best rated questionnaire, along with its simplified 11-item version (Quick DASH),⁷ not adapted to Spanish.⁴⁸ It is the single most transculturally adapted questionnaire.³⁷ The psychometric properties of the version by Hervás et al.,²⁵ produced in Spain, have been studied in depth, although its agreement values have not been analyzed. It is a long questionnaire, which, instead of being specific to the shoulder, addresses the entire upper limb, and would thus be generally preferable for use in research. In this respect, another option would be to use the ULFI, for which transcultural adaptation has been less rigorous²⁷ in the case of this scale, unlike the DASH, data are available on agreement but not on sensitivity. Only one of the three adaptations of the DASH, the version by Hervás et al.,²⁵ was carried out on patients with shoulder disabilities.

The psychometric properties of the SPADI and OSS have similarly been fully and rigorously analyzed, with the single exception of their agreement values. Both are very short. The OSS questionnaire is the most widely used in the UK,⁴⁹ and, at an international level, is the most widely adapted and validated general shoulder scale. Even so, it is relatively little used in scientific literature and is not suitable for patients with instability. Aside from being the most sensitive questionnaire, the SPADI is also the instrument that is most widely accepted by the scientific community;⁵⁰ it would be the first choice for both clinical practice and research,^{7,8} and would probably also be best for primary care.⁵¹

The SST and SDQ questionnaires are likewise short and applicable to any shoulder disability. The former is widely used in the USA⁷ but, because it only assesses functionality, should be associated with a pain scale. Furthermore, binary replies allow for few nuances or shades of difference. The latter instrument appears to have a notable ceiling effect, an aspect not analyzed in the Spanish version, in which validity was also not examined.

The WORC, WOSI and WUSPI questionnaires are specific for rotator cuff disease, shoulder instability and wheelchair users, respectively. The first two have been adapted to the Argentine population;^{32,33} both are somewhat long when it comes to their being applied. In the third,³⁴ the quality of the transcultural adaptation and validity assessment was poor, although reliability was good.

Based on the results of our review, within the questionnaires adapted to Spanish, we consider the SPADI to be the first choice for both clinical practice and research on shoulder pain, and the DASH to be generally preferable for use in research to address the entire upper limb, given that it is not specific to the shoulder. Among patients with rotator cuff disease, the SPADI would be preferable to the WORC because the

adaptation of the latter is of worse quality. For other specific situations, such as shoulder instability and wheelchair users, there are the WOSI and WUSPI questionnaires, respectively. These adaptations are of inferior quality to that of the SPADI questionnaire.

In the searches conducted, every effort was made to identify and retrieve all relevant questionnaires that had been adapted into Spanish. It has to be said, however, that even though the most commonly used databases (including local databases) were selected, some studies may not have been detected because some journals may not be indexed in these databases. Accordingly, this could be viewed as a limitation of our review. As strong points of our review, it should be stressed that both the rigor of the questionnaire-adaptation process and the psychometric properties studied were analyzed by two authors. Furthermore, the search process was so exhaustive as to render it extremely unlikely that any other published scales adapted to Spanish might have gone undetected.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. The paper has not been presented at any society or meeting.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

1. Linaker CH and Walker-Bone K. Shoulder disorders and occupation. *Best Pract Res Clin Rheum* 2015; 29: 405–423.
2. Barr KP. Rotator cuff disease. *Phys Med Rehabil Clin N Am* 2004; 15: 475–491.
3. Frau-Escobal P, Langa-Revert Y, Querol-Fuentes F, Mora-Amérigo E and Such-Sanz A. Shoulder musculoskeletal disorders in primary care. A cross-sectional study in a health care center of the Valencia Health Care Agency. *Fisioterapia* 2013; 35: 10–17.
4. Harvie P, Pollard TC, Chennagiri RJ and Carr AJ. The use of outcome scores in surgery of the shoulder. *J Bone Joint Surg* 2005; 87: 151–154.
5. Roe Y, Soberg HL, Bautz-Holter E and Ostensjo S. A systematic review of measures of shoulder pain and functioning using the International classification of functioning, disability and health (ICF). *BMC Musculoskelet Disord* 2013; 28: 73.
6. Wylie JD, Beckmann JT, Granger E and Tashjian RZ. Functional outcomes assessment in shoulder surgery. *World J Orthop* 2014; 5: 623–633.
7. Angst F, Schwyzer HK, Aeschlimann A, Simmen BR and Goldhahn J. Measures of adult shoulder function: Disabilities of the Arm, Shoulder, and Hand Questionnaire (DASH) and its short version

- (QuickDASH), Shoulder Pain and Disability Index (SPADI), American Shoulder and Elbow Surgeons (ASES) Society standardized shoulder assessment form, Constant (Murley) Score (CS), Simple Shoulder Test (SST), Oxford Shoulder Score (OSS), Shoulder Disability Questionnaire (SDQ), and Western Ontario Shoulder Instability Index (WOSI). *Arthritis Care Res* 2011; 63(Suppl 11): S174-S188.
8. Schmidt S, Ferrer M, González M, et al. Evaluation of shoulder-specific patient-reported outcome measures: a systematic and standardized comparison of a available evidence. *J Shoulder Elbow Surg* 2014; 23: 434-444.
 9. Muñoz J, Elosua P and Hambleton RK. Directrices para la traducción y adaptación de los tests: segunda edición. *Psicothema* 2013; 25: 151-157.
 10. Epstein J, Santo RM and Guillemin F. A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus. *J Clin Epidemiol* 2015; 68: 435-441.
 11. Instituto Cervantes. Informe: El español una lengua viva, 2015. <http://www.cervantes.es/imagenes/File/prensa/El%20espaol%20una%20lengua%20viva.pdf> (accessed 6 February 2016).
 12. Puga VO, Lopes AD and Costa LO. Assessment of cross-cultural adaptations and measurement properties of self-report outcome measures relevant to shoulder disability in Portuguese: a systematic review. *Rev Bras Fisioter* 2012; 16: 85-93.
 13. Murphy DR and López M. Neck and back pain specific outcome assessment questionnaires in the Spanish language: a systematic literature review. *Spine J* 2013; 13: 1667-1674.
 14. Altman DG. Practical statistics for medical research. London: Chapman & Hall; 1999.
 15. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007; 60: 34-42.
 16. Guillemin F, Bombardier C and Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol* 1993; 46: 1417-1432.
 17. Beaton DE, Bombardier C, Guillemin F and Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 2000; 25: 3186-3191.
 18. Wild D, Grove A, Martin M, 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 Health* 2005; 8: 94-104.
 19. Ramada-Rodilla JM, Serra-Pujadas C and Deldos-Clandet GL. Adaptación cultural y validación de cuestionarios de salud: revisión y recomendaciones metodológicas. *Salud Publica Mex* 2013; 55: 57-66.
 20. Lujan-Tangarife JA and Cardona Arias JA. Construcción y validación de escalas de medición en salud: revisión de propiedades psicométricas. *Arch Med* 2015; 11: 1.
 21. Munro B. *Statistical methods for health care research*. Philadelphia, PA: JB Lippincott, 2000.
 22. Michener LA and Leggin BG. A review of self-report scales for the assessment of functional limitation and disability of the shoulder. *J Hand Ther* 2001; 14: 68-76.
 23. Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM and 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* 2012; 21: 651-657.
 24. Rosales RS, Delgado EB and Díez de la Lastra-Bosch I. Evaluation of the Spanish version of the DASH and carpal tunnel syndrome health-related quality-of-life instruments: cross-cultural adaptation process and reliability. *J Hand Surg Am* 2002; 27: 334-343.
 25. Hervás MT, Navarro Collado MJ, et al. Spanish version of the DASH questionnaire. Cross-cultural adaptation, reliability, validity and responsiveness. *Med Clin (Barc)* 2006; 127: 441-447.
 26. Mulero-Portela AL, Colón-Santadía CL and Cruz-Gómez C. Cross-cultural adaptation of the Disability of Arm, Shoulder, and Hand questionnaire: Spanish for Puerto Rico version. *Int J Rehabil Res* 2009; 32: 287-293.
 27. Cuesta-Vargas AI and Gabel PC. Cross-cultural adaptation, reliability and validity of the Spanish version of the Upper Limb Functional Index. *Health Qual Life Outcomes* 2013; 11: 126.
 28. Membrilla-Mesa MD, Tejero-Fernández V, Cuesta-Vargas AI and Arroyo-Morales M. Validation and reliability of a Spanish version of Simple Shoulder Test (SST-Sp). *Qual Life Res* 2015; 24: 411-416.
 29. Membrilla-Mesa MD, Cuesta-Vargas AI, Pozuelo-Calvo R, Tejero-Fernández V, Martín-Martín L and Arroyo-Morales M. Shoulder Pain and Disability Index: cross cultural validation and evaluation of psychometric properties of the Spanish version. *Health Qual Life Outcomes* 2015; 13: 200.
 30. Torres-Lacomba M, Sánchez-Sánchez B, Prieto-Gómez V, et al. Spanish cultural adaptation and validation of the shoulder pain and disability index, and the Oxford shoulder score after breast cancer surgery. *Health Qual Life Outcomes* 2015; 13: 63.
 31. Álvarez-Nemegyei J, Puerto-Ceballos I, Guzmán-Hau W, Bassol-Perea A and Nuño-Gutiérrez BL. Development of a Spanish-language version of the Shoulder Disability Questionnaire. *J Clin Rheumatol* 2005; 11: 185-187.
 32. Arcuri F, Barclay F and Nacul I. Traducción, adaptación transcultural, validación y medición de propiedades de la versión en español del índice Western Ontario Rotator Cuff (WORC). *Artroscopia* 2015; 22: 56-60.
 33. Arcuri F, Nacul I and Barclay F. Traducción, adaptación transcultural, validación y medición de propiedades de la versión en español del índice Western Ontario Shoulder Instability (WOSI). *Artroscopia* 2015; 22: 125-133.
 34. Arroyo-Ajano R and González-Viejo MA. Validación al castellano del Wheelchair Users Shoulder Pain Index (WUSPI). *Rehabilitación* 2009; 43: 2-9.
 35. Arcuri F, Barclay F and Nacul I. Translation, cultural adaptation and validation of the Simple Shoulder Test to Spanish. *Orthop J Sports Med* 2014; 2(4 Suppl): 2325967114S00233.

36. Roddey TS, Cook KF, O'Malley KJ and Gartsman GM. The relationship among strength and mobility measures and self-report outcome scores in persons after rotator cuff repair surgery: impairment measures are not enough. *J Shoulder Elbow Surg* 2005; 14(1 Suppl S): 95–98.
37. St-Pierre C, Desmeules F, Dionne CE, Frémont P, MacDermid JC and Roy JS. Psychometric properties of self-reported questionnaires for the evaluation of symptoms and functional limitations in individuals with rotator cuff disorders: a systematic review. *Disabil Rehabil* 2016; 38: 103–122.
38. Huang H, Grant JA, Miller BS, Mirza FM and Gagnier JJ. A systematic review of the psychometric properties of patient-reported outcome instruments for use in patients with rotator cuff disease. *Am J Sports Med* 2015; 43: 2572–2582.
39. Röller AS, Mounts RA, DeLong JM and Hanypsiak BT. Outcome instruments for the shoulder. *Arthroscopy* 2013; 29: 955–964.
40. Badalamente M, Coffelt L, Elfar J, et al. Measurement scales in clinical research of the upper extremity, part 2: outcome measures in studies of the hand/wrist and shoulder/elbow. *J Hand Surg Am* 2013; 38: 407–412.
41. Roy JS and Esculier JF. Psychometric evidence for clinical outcome measures assessing shoulder disorders. *Phys Ther Rev* 2011; 16: 331–346.
42. Longo UG, Vasta S, Maffulli N and Denaro V. Scoring systems for the functional assessment of patients with rotator cuff pathology. *Sports Med Arthrosc* 2011; 19: 310–20.
43. Wright RW and Baumgarten KM. Shoulder outcomes measures. *J Am Acad Orthop Surg* 2010; 18: 436–444.
44. Roy JS, MacDermid JC and Woodhouse LJ. Measuring shoulder function: a systematic review of four questionnaires. *Arthritis Rheum* 2009; 61: 623–632.
45. Placzek JD, Lukens SC, Badalamenti S, et al. Shoulder outcome measures: a comparison of 6 functional tests. *Am J Sports Med* 2004; 32: 1270–1277.
46. Bot SD, Terwee CB, van der Windt DA, Bouter LM, Dekker J and de Vet HC. Clinimetric evaluation of shoulder disability questionnaires: a systematic review of the literature. *Ann Rheum Dis* 2004; 63: 335–341.
47. Kirkley A, Griffin S and Dainty K. Scoring systems for the functional assessment of the shoulder. *Arthroscopy* 2003; 19: 1109–1120.
48. Kennedy CA, Beaton DE, Smith P, et al. Measurement properties of the QuickDASH (disabilities of the arm, shoulder and hand) outcome measure and cross-cultural adaptations of the QuickDASH: a systematic review. *Qual Life Res* 2013; 22: 2509–2547.
49. Varghese M, Lamb J, Rambani R and Venkateswaran B. The use of shoulder scoring systems and outcome measures in the UK. *Ann R Coll Surg Engl* 2014; 96: 590–592.
50. Page MJ, McKenzie JE, Green SE, et al. Core domain and outcome measurement sets for shoulder pain trials are needed: systematic review of physical therapy trials. *J Clin Epidemiol* 2015; 68: 1270–1281.
51. Paul A, Lewis M, Shadforth MF, Croft PR, Van der Windt DA and Hay EM. A comparison of four shoulder-specific questionnaires in primary care. *Ann Rheum Dis* 2004; 63: 1293–1299.

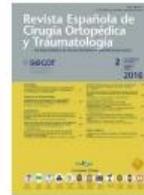
14.3. Anexo 3. Artículo de los cuestionarios autocumplimentados adaptados al español para la valoración funcional de pacientes con afecciones de rodilla.

Rev Esp Cir Ortop Traumatol. 2017;61(2):96-103



Revista Española de Cirugía
Ortopédica y Traumatología

www.elsevier.es/rot



ORIGINAL

Revisión sistemática de los cuestionarios autocumplimentados adaptados al español para la valoración funcional de pacientes con afecciones de rodilla



S. Gómez-Valero ^{a,*}, F. García-Pérez ^{a,b}, M.T. Flórez-García ^{a,b} y J.C. Miangolarra-Page ^b

^a Departamento de Medicina Física y Rehabilitación, Hospital Universitario Fundación Alcorcón, Alcorcón, Madrid, España

^b Departamento de Terapia Física, Terapia Ocupacional, Medicina Física y Rehabilitación, Universidad Rey Juan Carlos, Alcorcón, Madrid, España

Recibido el 30 de junio de 2016; aceptado el 18 de noviembre de 2016

Disponible en Internet el 31 de enero de 2017

PALABRAS CLAVE

Rodilla;
Evaluación de
resultados;
Cuestionario;
Adaptación
transcultural;
Versión española

Resumen

Objetivo: Realizar una revisión sistemática de cuestionarios autocumplimentados de valoración funcional para afecciones de rodilla adaptados al español analizando la calidad de la adaptación transcultural y las propiedades psicométricas.

Material y métodos: Se realizó una búsqueda en las principales bases de datos biomédicas para localizar escalas de valoración funcional de rodilla adaptadas al español, evaluando el proceso de adaptación y sus propiedades psicométricas.

Resultados: Se identificaron 10 escalas; 3 fueron para miembro inferior: 2 para cualquier tipo de afección (*Lower Limb Functional Index* [LLFI]; *Lower Extremity Functional Scale* [LEFS]) y una específica para artrosis (*Arthrose des Membres Inférieurs et Qualité de vie* [AMICAL]); otras 3 para patologías de rodilla y cadera (*Western Ontario and McMaster Universities Osteoarthritis* [WOMAC] *index*; *Osteoarthritis Knee and Hip Quality of Life* [OAKHQOL] *questionnaire*; *Hip and Knee Questionnaire*[HKQ]), y otras 4 para rodilla: 2 generales (*Knee Injury and Osteoarthritis Outcome Score* [KOOS]; *Knee Society Clinical Rating System* [KSS]) y 2 específicas (*Victorian Institute of Sport Assessment* [VISA-P] *questionnaire* para pacientes con tendinopatía rotuliana y *Kujala Score* para el dolor femoropatelar). El procedimiento de adaptación transcultural fue satisfactorio, aunque algo menos riguroso para los cuestionarios HKQ y LLFI. En ningún estudio se evaluaron todas las propiedades psicométricas. La fiabilidad se analizó en todos los casos, menos en el KSS. La validez se midió en todos los cuestionarios.

Conclusión: Las propiedades psicométricas analizadas fueron aceptables y similares a la versión original y a otras versiones adaptadas a otros idiomas.

© 2017 SECOT. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

* Autor para correspondencia.

Correo electrónico: saragova@gmail.com (S. Gómez-Valero).

<http://dx.doi.org/10.1016/j.recot.2016.11.002>

1888-4415/© 2017 SECOT. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

KEYWORDS

Knee;
Outcome assessment;
Questionnaire;
Cross-cultural
adaptation;
Spanish version

A systematic review of self-administered questionnaires for the functional assessment of patients with knee disabilities adapted into Spanish

Abstract

Objective: The aim of this study was to conduct a systematic review of self-administered knee-disability functional assessment questionnaires adapted to Spanish, analysing the quality of the transcultural adaptation procedure and the psychometric properties of the new version.

Material and methods: A search was conducted in the main biomedical databases to find knee-function assessment scales adapted into Spanish, in order to assess their questionnaire adaptation process as well as their psychometric properties.

Results: Ten scales were identified; 3 for lower limb: 2 for any type of pathologies (Lower Limb Functional Index [LLFI]; Lower Extremity Functional Scale [LEFS]) and 1 specific for arthrosis (Arthrosis des Membres Inférieurs et Qualité de vie [AMICAL]); Other 3 for knee and hip pathologies (Western Ontario and McMaster Universities Osteoarthritis [WOMAC] index; Osteoarthritis Knee and Hip Quality of Life [OAKHQOL] questionnaire; Hip and Knee Questionnaire [HKQ]), and other 4 for knee: 2 general scales (Knee Injury and Osteoarthritis Outcome Score [KOOS]; Knee Society Clinical Rating System [KSS]) and 2 specifics (Victorian Institute of Sport Assessment [VISA-P] questionnaire for patients with patellar tendinopathy and Kujala Score for patello-femoral pain). The transcultural adaptation procedure was satisfactory, albeit somewhat less rigorous for HKQ and LLFI. In no study were all psychometric properties assessed. Reliability was analyzed in all cases, except in KSS. Validity was measured in all questionnaires.

Conclusions: The transcultural adaptation procedure was satisfactory and the psychometric properties analysed were similar to both the original version and other versions adapted to other languages.

© 2017 SECOT. Published by Elsevier España, S.L.U. All rights reserved.

Introducción

En el manejo de las patologías de rodilla es esencial una correcta valoración funcional para poder evaluar los cambios clínicos y el efecto de los tratamientos. Esta valoración puede realizarse mediante escalas funcionales que informan del impacto de la enfermedad desde la perspectiva del paciente¹.

Existen gran número de escalas, habitualmente en forma de cuestionarios autocumplimentados, que miden la función y los síntomas en las diversas patologías de rodilla². Algunos son aplicables a cualquier afección del miembro inferior (incluyendo la rodilla). Otros están diseñados exclusivamente para la región de la rodilla, bien generales (para cualquier patología), o bien específicos.

Recientemente se han publicado varios artículos que revisan las escalas que se pueden utilizar en problemas de rodilla^{3,4}. No se han publicado revisiones similares de cuestionarios para patología de rodilla en español.

El español es el idioma más hablado en el mundo después del chino mandarín y está experimentando un aumento en el número de hablantes. En 2015, casi 470 millones de personas lo han usado como lengua nativa (6,7% de la población mundial)⁵. Pero la mayoría de estas escalas han sido desarrolladas en países anglosajones, lo que genera inconvenientes cuando se van a utilizar en países con lenguas o culturas diferentes. Tener que aplicar estas escalas en pacientes de una cultura diferente implica que se debe realizar una adecuada adaptación transcultural para que tenga el mismo sentido que la versión original^{6,7}. Además, se deben

analizar las propiedades psicométricas de la nueva versión para asegurarse de que se trata de una herramienta equivalente a la original⁸.

El objetivo de nuestra revisión ha sido efectuar una búsqueda bibliográfica exhaustiva sobre los cuestionarios autocumplimentados de rodilla adaptados al español, analizando tanto la calidad metodológica del proceso de adaptación transcultural como las propiedades psicométricas de la versión española.

Material y métodos**Estrategia de búsqueda**

Se ha realizado una revisión sistemática de artículos publicados sobre adaptación transcultural al español de cuestionarios autocumplimentados para valorar a pacientes con afecciones de rodilla y sobre su proceso de validación en nuestro idioma. Se efectuó una búsqueda bibliográfica, desde su creación hasta el 31 de mayo de 2016, en bases de datos electrónicas internacionales (Medline, Embase, CINAHL y Web of Science). Los términos utilizados y la estrategia de búsqueda en Medline fueron los siguientes: «outcome» or «questionnaire» or «score» or «functional scale» or «assessment tool» or «instruments» and «knee» or «disability» or «lower limb» and «Spanish» or «Spanish version» or «Spanish cultural adaptation» or «Spanish assessment» or «Spanish translation» or «cross-cultural adaptation» or «validation». Se realizó, también,

una búsqueda manual utilizando los nombres de las diferentes escalas de rodilla, como palabras clave, en las distintas bases de datos. Además se efectuó una búsqueda en la web, incluyendo Google Scholar, con el fin de poder abarcar otro tipo de publicaciones, incluyendo literatura gris. Finalmente se examinaron manualmente las referencias de los artículos obtenidos.

Criterios de selección

De cada artículo encontrado, 2 de los autores revisaron el título y el resumen. Si su lectura sugería que podía ser seleccionado, se realizó la lectura completa del artículo. Incluimos artículos, sin restricción idiomática, sobre estudios que describieran el proceso de adaptación transcultural al español de cuestionarios autoaplicados para realizar la valoración funcional de pacientes con afecciones de rodilla y que analizaran, también, las propiedades psicométricas de la nueva versión. Se excluyeron artículos que solo analizaran las propiedades de un cuestionario previamente adaptado, protocolos de investigación y resúmenes de congresos.

Análisis de la información

A partir de los artículos seleccionados, 2 autores estructuraron y analizaron los resultados del manuscrito. En caso de discrepancia se buscó el consenso y, cuando fue necesario, se recurrió a un tercer revisor. El análisis de la información incluyó:

Características de los participantes en el estudio. Se recogieron los datos de los pacientes que formaron parte del estudio: país y localidad, número total de pacientes incluidos, diagnóstico, edad y sexo. Se comprobó si incluían al menos 50 pacientes, que es el número mínimo recomendable para los estudios de adaptación transcultural^{9,10}.

Evaluación de la metodología utilizada para efectuar la adaptación transcultural. Se comprobó si coincidía con los 5 pasos habitualmente recomendados en la bibliografía internacional^{6,9,10} y si se siguieron correctamente. Los 5 pasos son: 1) traducción directa del cuestionario original al español (efectuado de forma independiente por al menos 2 traductores bilingües cuyo idioma nativo fuera el español); 2) síntesis de traducciones y solución de posibles discrepancias entre los traductores con algún miembro del equipo de investigación; 3) traducción inversa o retrotraducción (de la traducción consensuada en español al idioma original, de forma independiente, por al menos 2 traductores que no conocieran dicha versión); 4) revisión por un comité de expertos para desarrollar la versión prefinal asegurando la equivalencia semántica, idiomática, experiencial y conceptual de la escala; y 5) prueba piloto (pretest) del cuestionario prefinal con sujetos de habla española (se recomienda hacerlo, idealmente, con 30-40 personas, buscando ítems no contestados y posibles problemas de comprensión). En cada uno de los pasos se consideró si se había hecho correctamente o no, si no se había realizado, o si los autores no facilitaban la información necesaria.

Evaluación de las propiedades psicométricas de la versión española. Se analizaron los siguientes aspectos: fiabilidad, validez y sensibilidad al cambio^{6,10-13}. La fiabilidad evalúa la exactitud o precisión del instrumento e incluye

la consistencia interna, la reproducibilidad test-retest y la confiabilidad (*agreement*). La consistencia interna se calcula mediante el coeficiente alpha (α) de Cronbach. Cuando el instrumento está compuesto por subescalas hay que calcular la correlación entre los ítems que forman cada dominio con la escala total. El análisis factorial se aplica para determinar la dimensionalidad de los ítems. La reproducibilidad test-retest se calcula mediante el índice de correlación intraclass (ICC). La confiabilidad se valora mediante el *standard error of measurement* (SEM) y el *minimal detectable change* (MDC₉₀). La validez se mide mediante coeficientes de correlación como el de Pearson o Spearman, y la correlación puede ser directa, indirecta o ausente. Los efectos suelo y techo hacen referencia al porcentaje de sujetos que obtienen las puntuaciones más bajas y más altas posibles. No deben superar el 15% de los que contestan ya que, en caso contrario, se limita la validez de contenido de la escala debido a que los pacientes con valores extremos no pueden distinguirse entre sí. La sensibilidad es la capacidad del instrumento para detectar cambios clínicamente importantes a lo largo del tiempo. El cambio mínimo detectable se mide principalmente a través del *effect size* (ES) y del *standardized response mean* (SRM).

Evaluación de la aplicabilidad directa de las escalas adaptadas en un país hispanohablante para ser aplicadas en otro. Se analizó si las adaptaciones realizadas en países latinoamericanos necesitaban alguna modificación para poder ser utilizadas en España y viceversa.

Resultados

Mediante la búsqueda bibliográfica realizada identificamos 10 artículos relevantes¹⁴⁻²³. Se encontraron 10 cuestionarios autocomplimentados que habían sido traducidos y adaptados transculturalmente a la población española y de los que se habían estudiado las propiedades clínicas de la nueva versión. En todos los casos el cuestionario original se había desarrollado en inglés, excepto en 2: el cuestionario *Arthrose des Membres Inférieurs et Qualité de vie* (AMICAL)¹⁴ y el *Osteoarthritis Knee and Hip Quality of Life* (OAKHQOL) *questionnaire*¹⁹, desarrollados en francés. Había 3 cuestionarios aplicables a patologías del miembro inferior: 2 para cualquier tipo de afección (*Lower Limb Functional Index* [LLFI]²⁰ y *Lower Extremity Functional Scale* [LEFS]²¹) y uno específico para artrosis (AMICAL)¹⁴. Se encontraron otros 3 cuestionarios aplicables a patologías de rodilla y de cadera: *Western Ontario and McMaster Universities Osteoarthritis (WOMAC) index*¹⁷; OAKHQOL¹⁹ y *Hip and Knee Questionnaire* (HKQ)¹⁵. Había 2 cuestionarios generales de rodilla: *Knee Injury and Osteoarthritis Outcome Score* (KOOS)¹⁶ y *Knee Society Clinical Rating System* (KSS)¹⁸. Finalmente se hallaron 2 cuestionarios para situaciones específicas de rodilla: *Victorian Institute of Sport Assessment* (VISA-P) *questionnaire*²³ para la tendinopatía rotuliana y *Kujala Score*²³ para pacientes con dolor femoropatelar. Hay que señalar que nuestra búsqueda no encontró ningún cuestionario desarrollado originalmente en español.

En la *tabla 1* se pueden observar las características demográficas y clínicas de la población en la que se realizó cada uno de los artículos.

Tabla 1 Características demográficas y clínicas de la población de los estudios de adaptación transcultural

Escala	Autor (año de la adaptación)	País (población)	Número total de pacientes incluidos	Diagnóstico	Edad en años Media (DE)	Porcentaje de mujeres (%)
AMICAL ¹⁴	Espinosa-Cuervo et al. (2014)	México	100	Gonartrosis grado II-III radiológico de la clasificación de Kellgren y Lawrence	56,34 (±13)	67
HKQ ¹⁵	Castellet et al. (2014)	España (no especificado)	316	Gonartrosis con indicación de prótesis de rodilla antes y después de la cirugía	≥ 18	—
KOOS ¹⁶	Vaquero et al. (2012)	España (no especificado)	20	Defecto osteocondral y/o defecto osteocondral degenerativo	41,35 (±14) Rango: 24-70	70
WOMAC ¹⁷	Villanueva et al. (2004)	España (Sevilla)	73	Gonartrosis (al menos grado II de la clasificación de Kellgren y Lawrence)	58,71 Rango: 35-73	89
KSS ¹⁸	Ares et al. (2013)	España (no especificado)	316	Gonartrosis con indicación de prótesis de rodilla antes y a los 6 meses de la cirugía	—	—
OAKHQOL ¹⁹	González Sáenz de Tejada et al. (2011)	España (Canarias)	409	Gonartrosis	71,26 (±7,71)	63,3
LLF ²⁰	Cuesta-Vargas et al. (2014)	España (Málaga)	136	Patología crónica (>12 semanas) de extremidades inferiores	48 (±19)	54,4
LEFS ²¹	Cruz-Díaz et al. (2014)	España (Jaén)	132	Patología de un miembro inferior (un 1,5% diagnosticado de osteoartritis; 22% dolor muscular; 9,1% lesión ligamentosa; 23,5% lesión meniscal; 22,0% dolor femoropatelar, 7,6% fractura articular, muscular o de partes blandas)	27,11(±6,22)	44,69
VISA-P ²²	Hernández-Sánchez et al. (2011)	España (no especificado)	150	Tendinopatía rotuliana con tendinosis verificada por ecografía o resonancia magnética	23,5	7,5
Kujala ²³	Gil-Gómez et al. (2015)	España (no especificado)	130	Dolor femoropatelar reclutados en clubes deportivos y centros de fisioterapia	21,2 (±3,6)	71,54

AMICAL: Arthrose des Membres Inférieurs et Qualité de vie; DE: desviación estándar; HKQ: Hip and Knee Questionnaire; KOOS: Knee Injury and Osteoarthritis Outcome Score; KSS: Knee Society Clinical Rating System; LEFS: Lower Extremity Functional Scale; LLF: Lower Limb Functional Index; OAKHQOL: Osteoarthritis Knee and Hip Quality of Life questionnaire; VISA-P: Victorian Institute of Sport Assessment questionnaire; WOMAC: Western Ontario and McMaster Universities Osteoarthritis index.

Tabla 2 Evaluación de la metodología empleada para la adaptación transcultural de los cuestionarios

Escala	Traducción	Síntesis	Traducción inversa	Análisis por comité de expertos	Prueba piloto (número de pacientes)
AMICAL ¹⁴	+	+	+	+	+ (10)
HKQ ¹⁵	+	+	-	?	+ (20)
KOOS ¹⁶	+	+	+	+	+ (10)
WOMAC ¹⁷	+	+	-	+	+ (10)
KSS ¹⁸	+	+	+	+	+ (20)
OAKHQOL ¹⁹	+	+	+	+	+ (17)
LLFI ²⁰	+	+	-	+	0
LEFS ²¹	+	+	+	+	0
VISA-P ²²	+	+	+	+	+ (12)
Kujala ²³	+	+	+	+	+ (19)

+: correctamente hecho; ?: dudoso; -: Incorrectamente hecho o no hecho; 0: no se da información sobre si se hizo o no se hizo; AMICAL: *Arthrose des Membres Inférieurs et Qualité de Vie*; HKQ: *Hip and Knee Questionnaire*; KOOS: *Knee Injury and Osteoarthritis Outcome Score*; KSS: *Knee Society Clinical Rating System*; LEFS: *Lower Extremity Functional Scale*; LLFI: *Lower Limb Functional Index*; OAKHQOL: *Osteoarthritis Knee and Hip Quality of Life questionnaire*; VISA-P: *Victorian Institute of Sport Assessment questionnaire*; WOMAC: *Western Ontario and McMaster Universities Osteoarthritis Index*.

La escala adaptada, junto con las instrucciones para el paciente, se incluyen en la publicación en todos los casos salvo para 3 cuestionarios: KOOS¹⁶, WOMAC¹⁷ y KSS¹⁸. La mayoría de los artículos se han publicado muy recientemente. Una de las adaptaciones se publicó en 2015²³, 4 en 2014^{14,15,20,21}, una en 2013¹⁸ y otra en 2012¹⁶. La primera escala adaptada al español fue el WOMAC, en 2004¹⁷. Hay que señalar que esta escala se tradujo por primera vez al español en 1999²⁴, pero al no incluir el estudio de las propiedades psicométricas fue excluida de nuestra selección de artículos.

Todos los cuestionarios fueron validados en España, excepto uno (AMICAL)¹⁴ que se validó en México. En todas las adaptaciones, menos en una¹⁴, se incluyó un número de pacientes superior al mínimo recomendado de 50 para este tipo de estudios.

En general el proceso de adaptación transcultural de todas las escalas requirió modificaciones, aunque de poca importancia, respecto a la versión original.

Los principales resultados de la evaluación de la metodología utilizada para la adaptación transcultural se analizan en la tabla 2.

Seis cuestionarios cumplieron rigurosamente los 5 pasos de las recomendaciones de las guías internacionales^{14,16,18,19,22,23}. El resto de cuestionarios fueron menos estrictos en ese aspecto. La prueba piloto se realizó para 8 escalas^{14-19,22,23}, aunque ninguna se hizo con más de 30 sujetos.

Las principales propiedades psicométricas analizadas en las versiones españolas se muestran en la tabla 3. Ningún trabajo evaluó todas las propiedades métricas de la nueva versión. En general, la conclusión de los autores de cada una de las adaptaciones realizadas fue que las propiedades psicométricas evaluadas resultaban aceptables y comparables a las de las versiones originales y a otras versiones adaptadas a otros idiomas.

En lo referente a la fiabilidad, la consistencia interna se evaluó en todas las escalas menos en una¹⁸, la reproducibilidad test-retest en todas menos en 2^{15,18} y la confiabilidad solo se analizó en 2 escalas^{20,21}. Cuando se realizó el

test-retest solo 2 escalas^{14,17} indicaron el número de participantes, siendo una¹⁴ la que superó el número ideal de más de 29 pacientes recomendables para efectuar esta evaluación. La consistencia interna total en todas las escalas, menos en una¹⁹, fue > 0,7-0,8, valor suficiente para garantizar su fiabilidad. El ICC fue muy bueno en todos los casos. La confiabilidad fue analizada en 2 cuestionarios^{20,21}. Cruz-Díaz et al.²¹ analizaron el valor del MDC₉₅, que fue de 2,18, y el de SEM, que fue de 0,88. Cuesta-Vargas et al.²⁰ calcularon el valor de MDC₉₅ y el valor de SEM, que fueron de 7,12 y 3,12% respectivamente. La validez se evaluó en todos los cuestionarios. Esta variaba en función de los instrumentos de comparación pero, en general, fue adecuada.

Los efectos suelo y techo solo se analizaron en 4 cuestionarios^{15,18,20,23}, y en ninguno estaban presentes. La sensibilidad solo fue examinada en 2 escalas^{20,22}. En ellas se obtuvieron valores, para ES y SRM, que indicaban una sensibilidad grande en ambos casos.

Discusión

Dentro de los sistemas de evaluación de los resultados de los tratamientos un aspecto esencial es que el propio paciente, mediante cuestionarios, valore desde su perspectiva los resultados alcanzados. Las mediciones efectuadas por el clínico se correlacionan mal con la discapacidad percibida por el paciente y este aspecto es, sin embargo, uno de las más relevantes para él²⁵. Por ello los cuestionarios son herramientas complementarias del examen clínico. Se han desarrollado numerosos cuestionarios a nivel internacional para valorar el impacto autopercebido por pacientes con afecciones de la rodilla. Cada uno presenta ventajas y desventajas respecto a los demás y no está claro que ninguno sea superior al resto. El interés que despierta este asunto es creciente y se han publicado numerosas revisiones en los últimos años²⁻⁴. La mayoría de los cuestionarios se han desarrollado en países anglosajones. Es preferible adaptar una escala ya existente, comprobando que la nueva versión mantiene las propiedades psicométricas de la original,

Tabla 3 Principales propiedades psicométricas analizadas en los cuestionarios adaptados

Escala	Consistencia interna (α de Cronbach) total	Reproductibilidad test-retest: Número de participantes (tiempo entre las dos evaluaciones) ICC	Instrumento de comparación: validez (coeficiente de correlación) ^a	Efectos suelo/techo	Sensibilidad ES; SRM
AMICAL ¹⁴	0,946	65 (72 h) 0,979 (IC 95%: 0,961-0,991)	WOMAC: EVA ($r = 0,737$), capacidad funcional ($r = -0,831$), dolor ($r = -0,742$) y rigidez ($r = -0,649$), con $p < 0,05$ SF-36: actividad física ($r = 0,709$); rol físico ($r = 0,678$); salud mental ($r = 0,705$); salud general ($r = 0,506$); rol emocional ($r = 0,432$). En otros dominios, las correlaciones fueron bajas pero significativas: dolor ($r = 0,484$), función social ($r = 0,346$) y apoyo social ($r = 0,287$), con $p < 0,05$ WOMAC: dolor y funcionalidad ($r = 0,641$) y rigidez ($r = 0,533$) SF-36: funcionalidad ($r = 0,498$); rol físico ($r = 0,480$); salud mental ($r = 0,333$); salud general ($r = 0,316$); rol emocional ($r = 0,406$); social ($r = 0,405$); vitalidad ($r = 0,371$); dolor ($r = 0,537$); componente físico ($r = 0,457$); componente mental ($r = 0,360$) SF-36: correlación más alta con el dolor ($r > 0,7$)	-	-
HKQ ¹⁵	0,864	No estudiado		No/No	-
KOOS ¹⁷	> 0,7 (dolor: 0,88; síntomas: 0,66; AWD: 0,96; deportes/ocio: 0,91; calidad de vida: 0,88)	- > 0,8		-	-
WOMAC ¹⁸	Rango de 0,71 a 0,97 para el dolor y de 0,64 a 0,95 para función No estudiado	17 (72 h) 0,36 a 0,76 para dolor y 0,34 a 0,52 para función No estudiado	WOMAC: dolor ($r = 0,27$); función ($r = 0,77$)	-	-
OKRHOOL ¹⁹	Rango de 0,60 a 0,93	- Rango de 0,75 a 0,81	WOMAC: significativa con $p < 0,01$ en todos los coeficientes siendo cerca de 0,4 el resto de las dimensiones SF-36: correlación algo inferior, especialmente dentro de las dimensiones mentales, con $r = 0,4$ y $r = 0,3$ para los dominios de salud mental y rol emocional SF-36: moderadamente correlacionado con los dominios físicos (función física, $r = 0,58$; dolor, $r = 0,64$; salud general, $r = 0,41$; y CFR, $r = 0,59$) WOMAC: moderadamente correlacionado con la función ($r = -0,80$) y con el dolor ($r = -0,69$)	No/No	-
LLFI ²⁰	0,91	- 0,96 (rango de 0,93 a 0,97)	EDQ-5D: $r > 0,5$ WOMAC: fuerte ($r = 0,77$) EQ-5D-3L: Justo e inversamente correlacionados ($r = -0,62$)	-	-
LEFS ²¹	0,98	- 0,998 (IC 95%: 0,996-0,999)	SF-36: funcionalidad ($r = 0,896$); rol físico ($r = 0,800$); rol emocional ($r = 0,504$); social ($r = 0,764$); salud mental ($r = 0,744$); vitalidad ($r = 0,760$); dolor ($r = 0,903$); salud general ($r = 0,657$); componente físico ($r = 0,902$); componente mental ($r = 0,824$), con $p < 0,01$	No/No	2,3; 1,9
VISA-P ²²	0,885	- 0,994 (IC 95%: 0,992-0,996)	Kujala: $r = 0,897$; $p < 0,001$ Cincinnati: $r = 0,782$; $p < 0,001$ SF-36: $r > 0,6$; $p < 0,001$ VISA-P: $r = 0,7$; $p < 0,001$	-	1,14; 1,17
Kujala ²³	0,8	- 0,99		No/No	-

AMICAL: Arthrose des Membres Inférieurs et Qualité de vie; AWD: actividades de la vida diaria; CFR: Componente físico resumido; EQ-5D-3L: European Health Questionnaire 5 Dimensions 3 Levels; ES: effect size; EVA: escala visual analógica; HKQ: Hip and Knee Questionnaire; IC: intervalo de confianza; ICC: índice de correlación intraclase; KOOS: Knee Injury and Osteoarthritis Outcome Score; KSS: Knee Society Clinical Rating System; LEFS: Lower Limb Functional Index; LLFI: Lower Limb Functional Scale; LLI: Lower Limb Functional Index; OKRHOOL: Osteoarthritis Knee and Hip Quality of Life questionnaire; SF-36: Short Form 36 Health Survey; SRM: standardized response mean; VISA-P: Victorian Institute of Sport Assessment questionnaire; WOMAC: Western Ontario and MacMaster Universities Osteoarthritis index.

^a Los valores negativos indican correlación inversa.

que crear otra nueva. Esto último supone un mayor esfuerzo económico y de tiempo e incrementa la diversidad de cuestionarios. Usar una misma escala facilita las comparaciones entre diferentes poblaciones.

El objetivo de nuestro trabajo ha sido revisar sistemáticamente los cuestionarios autocumplimentados de rodilla adaptados al español analizando tanto la calidad metodológica del proceso de adaptación transcultural como las propiedades psicométricas de la nueva versión obtenida. Hemos encontrado 10 artículos¹⁴⁻²³ sobre 10 cuestionarios, 5 de ellos publicados en los últimos 2 años^{14,15,20,21,22}.

Tres cuestionarios eran aplicables para patologías del miembro inferior: 2 para cualquier tipo de afección (LLFI²⁰ y LEFS²¹) y uno para artrosis (AMICAL)¹⁴. Tres para patologías de rodilla y cadera: WOMAC¹⁷, OAKHQOL¹⁹ y HKQ¹⁵. Cuatro para patologías de rodilla: 2 generales (KOOS¹⁶ y KSS¹⁸) y 2 específicos (VISA-P²² para la tendinopatía rotuliana y Kujala²³ para pacientes con dolor femoropatelar). De todos se dispone, también, de adaptaciones transculturales a otros idiomas.

No hay todavía un claro consenso internacional sobre el modo óptimo de realizar la adaptación transcultural. Si parece haber acuerdo en que la traducción inversa o retrotraducción y la prueba piloto (pretest) resultan esenciales⁶. Los criterios reconocidos internacionalmente referentes al proceso de adaptación al español fueron seguidos con el máximo rigor para 6 cuestionarios^{14,16,18,19,22,23}, pero también los restantes pueden ser considerados metodológicamente válidos. No obstante, el proceso fue algo menos riguroso para los cuestionarios HKQ y LLFI. En las diferentes naciones o poblaciones hispanohablantes pueden existir palabras de uso exclusivo o con significado diferente a los países o grupos culturales. No obstante, en general, no hay diferencias culturales tan marcadas, que impidan el uso de estos cuestionarios en los diferentes países que hablan español. En el cuestionario AMICAL, adaptado en México¹⁴, la introducción de pequeños cambios en algunas palabras usadas permitiría emplearlos en España. Por ejemplo, se podría sustituir «auto» por «coche» o «pesero» por «autobús».

Las propiedades psicométricas analizadas son adecuadas en todas las escalas y, en general, similares a las de la versión original y a la de otras versiones de la escala adaptadas a otros idiomas. Para seleccionar qué cuestionario utilizar en una situación concreta hay que considerar, por ello, criterios prácticos (tiempo necesario para cumplimentarlo y puntuarlo) y utilidad para determinadas patologías.

Una posible limitación de nuestro estudio es que para la valoración no hemos seguido, por su complejidad, ni la lista de verificación propuesta por el grupo COSMIN¹³ ni la herramienta llamada EMPRO²⁴, recomendadas para comprobar la adecuada validación de las propiedades psicométricas de un cuestionario. Como puntos fuertes de nuestra revisión podríamos destacar, en primer lugar, que tanto el rigor del proceso de adaptación del cuestionario como las propiedades psicométricas estudiadas fueron analizadas por 2 autores. Además, el proceso de búsqueda fue muy exhaustivo, por lo que resulta muy improbable que existan publicadas otras escalas adaptadas al español que hayan podido no ser encontradas.

En conclusión, el proceso de adaptación transcultural al español fue satisfactorio en los 10 cuestionarios y las propiedades psicométricas de las nuevas versiones fueron

aceptables y similares a las de los cuestionarios originales y a otras versiones adaptadas a otros idiomas.

Nivel de evidencia

Nivel de evidencia II.

Responsabilidades éticas

Protección de personas y animales. Los autores declaran que para esta investigación no se han realizado experimentos en seres humanos ni en animales.

Confidencialidad de los datos. Los autores declaran que en este artículo no aparecen datos de pacientes.

Derecho a la privacidad y consentimiento informado. Los autores declaran que en este artículo no aparecen datos de pacientes.

Conflicto de intereses

Los autores declaran no tener conflicto de intereses para los fines de este estudio.

Bibliografía

1. Beaton DE, Bombardier C, Gullieimin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976)*. 2000;25:3186-91.
2. Collins NJ, Misra D, Felson DT, Crossley KM, Roos EM. Measures of knee function: International Knee Documentation Committee (IKDC) Subjective Knee Evaluation Form, Knee Injury and Osteoarthritis Outcome Score (KOOS), Knee Injury and Osteoarthritis Outcome Score Physical Function Short Form (KOOS-PS), Knee Outcome Survey Activities of Daily Living Scale (KOS-ADL), Lysholm Knee Scoring Scale, Oxford Knee Score (OKS), Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), Activity Rating Scale (ARS), and Tegner Activity Score (TAS). *Arthritis Care Res (Hoboken)*. 2011;63 Suppl. 11:S208-28.
3. Rolfsen O, Ereslan Chenok K, Bohm E, Lübbecke A, Dentssen G, Dunn J, et al. Patient-reported outcome measures in arthroplasty registries. *Acta Orthop*. 2016;87 Suppl 1:3-8.
4. White DK, Master H. Patient-reported measures of physical function in knee osteoarthritis. *Rheum Dis Clin North Am*. 2016;42:239-52. <http://dx.doi.org/10.1016/j.rdc.2016.01.005>.
5. Instituto Cervantes. El español: una lengua viva; 2015 [consultado 6 Feb 2016]. Disponible en: <http://www.cervantes.es/imagenes/File/prensa/EI%20espaol%20una%20lengua%20viva.pdf>
6. Gullieimin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol*. 1993;46:1417-32.
7. Epstein J, Santo RM, Gullieimin F. A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus. *J Clin Epidemiol*. 2015;68:435-41. <http://dx.doi.org/10.1016/j.jclinepi.2014.11.021>.
8. Scholtes VA, Terwee CB, Poolman RW. What makes a measurement instrument valid and reliable? *Injury*. 2011;42:236-40.
9. Altman DG. *Practical statistics for medical research*. London: Chapman & Hall; 1999.

10. 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. *J Clin Epidemiol*. 2007;60:34–42.
11. 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 Health*. 2005;8:94–104.
12. Ramada-Rodilla JM, Serra-Pujadas C, Delcós-Cianchet GL. Adaptación cultural y validación de cuestionarios de salud: revisión y recomendaciones metodológicas. *Salud Publica Mex*. 2013;55:57–66.
13. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: A scoring system for the COSMIN checklist. *Qual Life Res*. 2012;21:651–7, <http://dx.doi.org/10.1007/s11136-011-9960-1>.
14. Espinosa-Cuervo G, Guillermin F, Rat AC, Duarte-Salazar C, Alemán-Hernández SI, Vergara-Álvarez Y, et al. Transculturalization and validation of a Spanish translation of the specific lower limb osteoarthritis and quality of life questionnaire AMICAL: Arthrose des Membres Inférieurs et Qualité de vie AMIQUAL. *Reumatol Clin*. 2014;10:241–7, <http://dx.doi.org/10.1016/j.reuma.2013.07.008>.
15. Castellet E, Ares O, Celaya F, Valenti-Azcárate A, Salvador A, Torres A, et al., SEROD group. Transcultural adaptation and validation of the Hip and Knee questionnaire Into Spanish. *Health Qual Life Outcomes*. 2014;12:76, <http://dx.doi.org/10.1186/1477-7525-12-76>.
16. Vaquero J, Longo UG, Forriol F, Martinelli N, Vethencourt R, Denaro V. Reliability, validity and responsiveness of the Spanish version of the Knee Injury and Osteoarthritis Outcome Score (KOOS) in patients with chondral lesion of the knee. *Knee Surg Sports Traumatol Arthrosc*. 2014;22:104–8, <http://dx.doi.org/10.1007/s00167-012-2290-1>.
17. Villanueva I, Guzman MM, Toyos FJ, Ariza-Ariza R, Navarro F. Relative efficiency and validity properties of a visual analogue vs a categorical scaled version of the Western Ontario and McMaster Universities Osteoarthritis (WOMAC) index: Spanish versions. *Osteoarthritis Cartilage*. 2004;12:225–31.
18. Ares O, Castellet E, Maculé F, León V, Montañez E, Freire A, et al. Translation and validation of 'The Knee Society Clinical Rating System' into Spanish. *Knee Surg Sports Traumatol Arthrosc*. 2013;21:2618–24, <http://dx.doi.org/10.1007/s00167-013-2412-4>.
19. González Sáenz de Tejada M, Escobar A, Herdman M, Herrera C, García L, Sarasqueta C. Adaptation and validation of the Osteoarthritis Knee and Hip Quality of Life (OAKHQoL) questionnaire for use in patients with osteoarthritis in Spain. *Clin Rheumatol*. 2011;30:1563–75, <http://dx.doi.org/10.1007/s10067-011-1855-6>.
20. Cuesta-Vargas AI, Gabel CP, Bennett P. Cross cultural adaptation and validation of a Spanish version of the Lower Limb Functional Index. *Health Qual Life Outcomes*. 2014;12:75, <http://dx.doi.org/10.1186/1477-7525-12-75>.
21. Cruz-Díaz D, Lomas-Vega R, Osuna-Pérez MC, Hita-Contreras F, Fernández AD, Martínez-Amat A. The Spanish lower extremity functional scale: A reliable, valid and responsive questionnaire to assess musculoskeletal disorders in the lower extremity. *Disabil Rehabil*. 2014;36:2005–11, <http://dx.doi.org/10.3109/09638288.2014.890673>.
22. Hernández-Sánchez S, Hidalgo MD, Gómez A. Cross-cultural adaptation of VISA-P score for patellar tendinopathy in Spanish population. *J Orthop Sports Phys Ther*. 2011;41:581–91, <http://dx.doi.org/10.2519/jospt.2011.3613>.
23. Gil-Gómez J, Pecos-Martín D, Kujala UM, Martínez-Mertero P, Montañez-Aguilera FJ, Romero-Franco N, et al. Validation and cultural adaptation of Kujala Score in Spanish. *Knee Surg Sports Traumatol Arthrosc*. 2015, <http://dx.doi.org/10.1007/s00167-015-3521-z>.
24. Battie-Guada E, Esteve-Vives J, Piera MC, Hargreaves R, Cuts J. Traducción y adaptación al español del cuestionario WOMAC específico para artrosis de rodilla y cadera. *Rev Esp Reumatol*. 1999;26:38–45.
25. Anderson JG, Wixson RL, Tsai D, Stulberg SD, Chang RW. Functional outcome and patient satisfaction in total knee patients over the age of 75. *J Arthroplasty*. 1996;11:831–40.
26. Valderas JM, Ferrer M, Mendivil J, Garín O, Rajmil L, Herdman M, et al., Scientific Committee on Patient-Reported Outcomes of the IRYSS Network. Development of EMPRO: A tool for the standardized assessment of patient-reported outcome measures. *Value Health*. 2008;11:700–8.

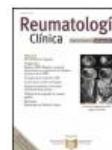
14.4. Anexo 4. Artículo sobre los cuestionarios autocumplimentados adaptados al español para la valoración funcional de pacientes con patología de cadera.

Reumatol Clin. 2018;14(6):372-378



Reumatología Clínica

www.reumatologiaclinica.org



Original

Revisión sistemática de los cuestionarios autocumplimentados adaptados al español para la valoración funcional de pacientes con patología de cadera



Sara Gómez-Valero^{a,*}, Fernando García-Pérez^{a,b}, Mariano Tomás Flórez-García^{a,b} y Juan Carlos Miangolarra-Page^b

^a Departamento de Medicina Física y Rehabilitación, Hospital Universitario Fundación Alcorcón, Alcorcón, Madrid, España

^b Departamento de Terapia Física, Terapia Ocupacional, Medicina Física y Rehabilitación, Universidad Rey Juan Carlos, Alcorcón, Madrid, España

INFORMACIÓN DEL ARTÍCULO

Historia del artículo:
Recibido el 5 de septiembre de 2016
Aceptado el 26 de enero de 2017
On-line el 17 de marzo de 2017

Palabras clave:
Cadera
Evaluación de resultados
Cuestionario
Adaptación transcultural
Versión española

RESUMEN

Objetivo: Realizar una revisión sistemática de la calidad de los procedimientos de adaptación y de las propiedades psicométricas de los cuestionarios autoadministrados de valoración funcional de cadera adaptados a la población española.

Material y método: Se realizó una búsqueda en las bases de datos MEDLINE, EMBASE, CINAHL y Web of Science (desde su inicio hasta junio de 2016) para localizar todas las escalas adaptadas al español y analizar las diferentes fases del proceso de adaptación y sus propiedades psicométricas.

Resultados: Se identificaron 8 escalas que se agruparon en 3 apartados, según el tipo de patologías en las que se pueden utilizar, a) todo el miembro inferior: *Lower Limb Functional Index* (LLFI), *Lower Extremity Functional Scale* (LEFS) y *Arthrose des Membres Inférieurs et Qualité de vie* (AMICAL); b) rodilla y/o cadera: *Western Ontario and McMaster Universities Osteoarthritis* (WOMAC) *index*, *Osteoarthritis Knee and Hip Quality of Life* (OAKHQOL) y *Hip and Knee Questionnaire* (HKQ), y c) exclusivas de cadera: *Hip Outcome Score* (HOS) e *International Hip Outcome Tool-33* (iHOT-33). El procedimiento de adaptación transcultural fue aceptable en todos los casos, aunque algo menos riguroso para las escalas HKQ y LLFI. Ningún estudio evaluó todas las propiedades psicométricas.

Conclusión: Disponemos actualmente de 8 cuestionarios de valoración funcional de cadera adaptados al español con aceptables características psicométricas. Podemos medir la repercusión percibida por el paciente de su patología de cadera seleccionando, entre las diferentes opciones, aquellas alternativas que mejor se adapten a nuestros objetivos, ya sean clínicos o de investigación.

© 2017 Elsevier España, S.L.U. y Sociedad Española de Reumatología y Colegio Mexicano de Reumatología. Todos los derechos reservados.

Systematic review of cross-cultural adaptations of hip-specific patient-reported outcome measures in Spanish

ABSTRACT

Objective: The aim of this study was to conduct a systematic review of the quality of the transcultural adaptation procedure and the clinimetric properties of the self-administered hip-disability functional assessment questionnaires adapted for the Spanish population.

Material and method: We searched the MEDLINE, EMBASE, CINAHL and Web of Science databases (from inception until June 2016) to locate all the scales adapted to Spanish and to analyze the different phases of the adaptation process and its psychometric properties.

Keywords:
Hip
Evaluation of results
Questionnaire
Cross-cultural adaptation
Spanish version

* Autor para correspondencia.
Correo electrónico: saragova@gmail.com (S. Gómez-Valero).

<https://doi.org/10.1016/j.reuma.2017.01.014>

1699-258X/© 2017 Elsevier España, S.L.U. y Sociedad Española de Reumatología y Colegio Mexicano de Reumatología. Todos los derechos reservados.

Results: Eight scales were identified, and were grouped into three sections, according to the type of diseases in which they can be used: a) lower limb: *Lower Limb Functional Index* (LLFI), *Lower Extremity Functional Scale* (LEFS) and *Arthrosis of Membres Inférieures et Qualité de vie* (AMICAL); b) knee and/or hip: *Western Ontario and McMaster Universities Osteoarthritis* (WOMAC) index, *Osteoarthritis Knee and Hip Quality of Life* (OAKHQOL) and *Hip and Knee Questionnaire* (HKQ); and c) specific for hip: *Hip Outcome Score* (HOS) and *International Hip Outcome Tool-33* (iHOT-33). The transcultural adaptation procedure was satisfactory in all cases, albeit somewhat less rigorous for the HKQ and LLFI than for the remaining questionnaires. No study evaluated all the psychometric properties.

Conclusion: We currently have 8 hip-disability functional assessment questionnaires adapted to Spanish with satisfactory psychometric properties. We can measure the patient's perceived impact of his or her hip disease by selecting, among the different options, those alternatives that best fit our clinical or research objectives.

© 2017 Elsevier España, S.L.U. and Sociedad Española de Reumatología y Colegio Mexicano de Reumatología. All rights reserved.

Introducción

En el manejo de las patologías de cadera es esencial una correcta valoración para poder evaluar la evolución de los síntomas y de la limitación funcional y conocer la eficacia de los distintos tratamientos. Esta valoración puede realizarse mediante diversos métodos. En la práctica clínica, cada vez se da más importancia a la medición de la salud y de la calidad de vida, siendo numerosos los instrumentos desarrollados para medir ambos parámetros¹. Las escalas de valoración funcional proporcionan información del impacto de la enfermedad desde el punto de vista del paciente, y son un complemento de la exploración física (movilidad, fuerza, etc.) y de los resultados obtenidos en las pruebas de imagen². Son numerosos los cuestionarios desarrollados a nivel internacional para valorar el impacto percibido por los pacientes con patología de cadera. La mayoría de estos cuestionarios han sido desarrollados en países anglosajones, lo que genera inconvenientes cuando se van a utilizar en países con lenguas o culturas diferentes. Para aplicarlos en pacientes de otra cultura es preciso realizar una adaptación transcultural adecuada^{3,4} (proceso más complejo que la simple traducción) y la validación de la escala⁵, es decir, debe establecerse su fiabilidad y validez en los nuevos grupos lingüísticos y culturales en los que se aplique⁶. Tras el chino mandarín, el español es el segundo idioma más hablado en el mundo, con una amplia extensión geográfica⁷, tanto en países que lo emplean como lengua oficial como en aquellos donde es utilizado por inmigrantes que residen en países con otras lenguas maternas.

En 2013 se han revisado los cuestionarios adaptados al español para pacientes con dolor cervical y lumbar⁸, pero no hemos encontrado revisiones similares para patología de la cadera.

El objetivo de este artículo es conocer los cuestionarios autocomplimentados de cadera adaptados a la población española analizando, además, la calidad metodológica del proceso de adaptación transcultural y las propiedades psicométricas de la nueva versión obtenida.

Material y métodos

Estrategia de búsqueda

Se ha realizado una revisión sistemática de los artículos publicados sobre todos los cuestionarios de valoración funcional autocomplimentados de cadera adaptados y validados al español. Se efectuó la búsqueda bibliográfica, desde su creación hasta junio de 2016, en las bases de datos electrónicas MEDLINE (1966-2016), EMBASE (1974-2016), CINAHL (1977-2016) y Web of Science (1900-2016). Los términos utilizados y la estrategia de búsqueda en MEDLINE fueron los siguientes: «Outcome» or «Questionnaire» or «Assessment» or «Instruments» and «Hip» or «Lower limb»

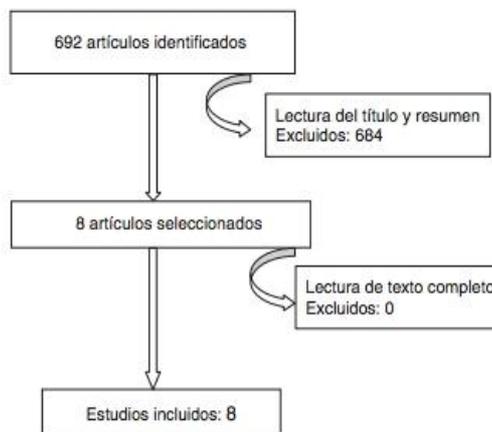


Figura 1. Diagrama de flujo que muestra la selección de los estudios incluidos.

and «Spanish» or «Spanish version» or «Spanish validation» or «Spanish translation» or «Cross-cultural adaptation» or «Cross-cultural validation». Se realizó, también, una búsqueda manual utilizando los nombres de las diferentes escalas de cadera, como palabras clave, en las distintas bases de datos. Además se efectuó una búsqueda en la web, incluyendo Google Scholar, con el fin de poder abarcar otro tipo de publicaciones. Finalmente se examinaron manualmente las referencias de los artículos obtenidos. La figura 1 muestra la selección de los estudios incluidos.

Criterios de selección

De cada artículo encontrado, 2 de los autores revisaron el título y el resumen. Si estos sugerían que podía ser seleccionado, se realizó la lectura completa del artículo. Incluimos artículos, sin restricción idiomática, sobre estudios que describieran el proceso de adaptación transcultural al español de cuestionarios autoaplicados para realizar la valoración funcional de pacientes con patología de cadera que analizaran, también, las propiedades psicométricas de la nueva versión. Se excluyeron artículos en los que solo se analizaran las propiedades de un cuestionario previamente adaptado, protocolos de investigación o resúmenes de congresos.

Análisis de la información

A partir de los artículos seleccionados, 2 autores estructuraron y analizaron los resultados del manuscrito. En caso de discrepancia

se buscó el consenso y, cuando fue necesario, se recurrió a un tercer revisor. El análisis de la información incluyó:

- Características de los participantes.** Se recogieron los datos de los pacientes que formaron parte del estudio: país y localidad, número total de pacientes incluidos, diagnóstico, edad y sexo. Se comprobó si incluían al menos 50 pacientes, que es el número mínimo recomendable para los estudios de adaptación transcultural^{4,5}.
- Evaluación de la adaptación transcultural.** Se comprobó si coincidía con los 5 pasos habitualmente recomendados en la bibliografía internacional^{5,9,10} y si se siguieron correctamente. Los cinco pasos son: 1) traducción directa del cuestionario original al español (efectuado de forma independiente por al menos 2 traductores bilingües); 2) síntesis de traducciones y solución de posibles discrepancias; 3) traducción inversa o retro-traducción (de la traducción consensuada en español al idioma original, por al menos 2 traductores independientes que no conocieran dicha versión); 4) revisión por un comité de expertos para desarrollar la versión pre-final asegurando la equivalencia semántica, idiomática y conceptual de la escala; y 5) prueba piloto (pre-test) del cuestionario pre-final con sujetos de habla española (se recomienda hacerlo con 30-40 personas), buscando ítems no contestados y posibles problemas de comprensión. En cada uno de los pasos se consideró como positivo (+) cuando el procedimiento fue realizado correctamente, dudoso (?) cuando la descripción no era clara, negativo (-) cuando se había realizado correctamente pero con una insuficiente cantidad de traductores/pacientes, o, finalmente, cero (0) cuando no se disponía de información suficiente para evaluar cada paso.
- Evaluación de las propiedades psicométricas.** Se analizaron los siguientes aspectos: fiabilidad, validez y sensibilidad a los cambios clínicos^{5,10-14}. La fiabilidad se evaluó mediante 3 parámetros: a) la consistencia interna, para conocer la estabilidad de las puntuaciones entre los diferentes elementos que componen una escala —medida de homogeneidad— a través de la prueba alfa de Cronbach, considerándose suficiente valores de alfa mayores que 0,7-0,8; b) la reproducibilidad test-retest o intervalo de tiempo adecuado —estabilidad temporal— medida con el índice de correlación intraclass (ICC) considerándose adecuados valores mayores de 0,8, y c) el grado de acuerdo que se valora mediante el error estándar de la medida (EEM) y el mínimo cambio detectable (MCD).
La validez se midió mediante 2 parámetros: a) los coeficientes de correlación de Pearson o Spearman, cuyos valores oscilan entre -1 y +1. El valor 0 indica que no existe asociación lineal entre las 2 variables de estudio; b) los efectos suelo y techo que aparecen cuando más del 15% de los encuestados logran la puntuación total más alta (techo) o más baja (suelo) posible.
La sensibilidad a los cambios clínicos es la capacidad del instrumento para detectar cambios clínicamente importantes a lo largo del tiempo. Se mide principalmente a través del tamaño del efecto (TE) y de la respuesta media estandarizada (RME).
- Evaluación de la aplicabilidad en diferentes países hispanohablantes.** Se analizó si las adaptaciones realizadas en países latinoamericanos necesitaban alguna modificación para poder utilizarse en España y viceversa.

Resultados

Se identificaron 8 artículos relevantes¹⁵⁻²², que describían 8 cuestionarios autocumplimentados adaptados transculturalmente a la población española y de los que se habían estudiado las propiedades clínicas de la nueva versión. Había 3 cuestionarios aplicables a cualquier patología del miembro inferior:

Lower Limb Functional Index (LLFI)¹⁵; *Lower Extremity Functional Scale* (LEFS)¹⁶ y el cuestionario *Arthrose des Membres Inférieurs et Qualité de vie* (AMICAL)¹⁷. Tres cuestionarios eran aplicables a patologías de rodilla y de cadera: *Western Ontario and McMaster Universities Osteoarthritis* (WOMAC) *index*¹⁸; *Osteoarthritis Knee and Hip Quality of Life* (OAKHQOL)¹⁹ y *Hip and Knee Questionnaire* (HKQ)²⁰. Había, por último, 2 cuestionarios específicos de cadera: *Hip Outcome Score* (HOS)²¹ y el *International Hip Outcome Tool-33* (iHOT-33)²².

En todos los casos el cuestionario original se había desarrollado en inglés, excepto en 2: AMICAL¹⁷ y OAKHQOL¹⁹, diseñados en francés. Nuestra búsqueda no localizó ningún cuestionario desarrollado originalmente en español.

En la *tabla 1* se muestran las características demográficas y clínicas de la población en la que se realizó cada uno de los estudios. La escala adaptada, junto con las instrucciones para el paciente, se incluye en la publicación en todos los casos salvo para 2 cuestionarios: WOMAC¹⁸ e iHOT-33²².

La mayoría de los artículos se han publicado muy recientemente. Una de las adaptaciones se publicó en 2015²² y 5 en 2014^{15-17,19,21}. La primera escala adaptada al español fue el WOMAC, en 2004¹⁸. Esta escala se adaptó por primera vez al español en 1999²³, pero al no incluir el estudio de las propiedades psicométricas fue excluida de nuestra selección de artículos.

Todos los cuestionarios se validaron en España, excepto uno¹⁷ que se validó en México. En todas las adaptaciones se incluyó un número de pacientes superior al mínimo recomendado de 50 para este tipo de estudios.

En general, el proceso de adaptación transcultural de todas las escalas requirió modificaciones, aunque de poca importancia, respecto a la versión original. Los principales resultados de la evaluación de la metodología utilizada para la adaptación transcultural se analizan en la *tabla 2*. Cinco cuestionarios cumplieron rigurosamente los 5 pasos de las recomendaciones de las guías internacionales^{16,17,19,21,22}. El resto fueron menos estrictos en ese aspecto. La prueba piloto se realizó en 6 escalas¹⁷⁻²², aunque solo 2^{21,22} la hicieron con 30 pacientes.

Las principales propiedades psicométricas analizadas en las versiones españolas se muestran en la *tabla 3*. Ningún trabajo evaluó todas las propiedades métricas de la nueva versión. En general, la conclusión de los autores de cada una de las adaptaciones realizadas fue que las propiedades psicométricas evaluadas resultaban aceptables y comparables a las de las versiones originales y a otras versiones adaptadas a otros idiomas.

En lo referente a la fiabilidad, la consistencia interna se evaluó en todas las escalas; en cuanto a la reproducibilidad test-retest, solo se especifica en 4 de ellas el tiempo entre escalas^{17,18,21,22} y de esas, solo en dos^{17,18} consta el número de pacientes a los que se aplicó, siendo una¹⁷ la que superó el número ideal de más de 29 pacientes recomendables para efectuar esta evaluación; el grado de acuerdo solo se analizó en 2 escalas^{21,22}.

La consistencia interna total en todas las escalas fue superior a 0,7-0,8, valor suficiente para garantizar su fiabilidad. El ICC fue muy bueno en todos los casos. El grado de acuerdo fue analizado en 2 cuestionarios^{21,22}. Seijas et al.²¹ analizaron el valor de MCD, que fue de 13,7 para la subescala de actividades de la vida diaria (AVD) y de 22,8 para la subescala de deporte, y el EEM fue de $\pm 5,1$ en AVD y de $\pm 8,5$ para deporte. Ruiz-Ibán et al.²² calcularon el valor de MCD y de EEM de 12,5 y 4,66, respectivamente.

La validez se evaluó en todos los cuestionarios. Esta variaba en función de los instrumentos de comparación pero, en general, fue adecuada. Los efectos suelo y techo solo se analizaron en 4 cuestionarios^{16,20-22}, estando presente el efecto techo en uno²¹ y el efecto suelo en otro²².

La sensibilidad se examinó solo en 3 escalas^{16,21,22}. En 2 de ellas únicamente se evaluó el RME^{21,22}, y solo en una¹⁶ se obtuvieron

Tabla 1
Características demográficas y clínicas de la población de los estudios de adaptación transcultural

Escala	Autor, año de la adaptación	País (población)	Número total de pacientes incluidos	Diagnóstico	Edad en años	Porcentaje de mujeres (%)
LLFI ¹⁵	Cuesta-Vargas et al., 2014	España (Málaga)	136	Patología crónica (> 12 semanas) de extremidades inferiores	Media: 48 (DE ± 19)	54,4
LEFS ¹⁶	Cruz-Díaz et al., 2014	España (Jaén)	132	Patología de un miembro inferior (un 1,5% diagnosticado de osteoartritis; 22% dolor muscular; 9,1% lesión ligamentosa; 23,5% lesión meniscal; 22,0% dolor femoropatelar; 7,6% fractura articular, muscular o de partes blandas).	Media: 27,11 (DE ± 6,22)	44,7
AMICAL ¹⁷	Espinosa-Cuervo et al., 2014	México	100	Coxartrosis grado II-III radiológico de la clasificación de Kellgren y Lawrence	Media: 56,34 (DE ± 13)	67
WOMAC ¹⁸	Villanueva et al., 2004	España (Sevilla)	73	Coxartrosis (al menos grado II de la clasificación de Kellgren y Lawrence)	Media: 58,71 Rango: 35-73	89
OAKHQOL ¹⁹	González Sáenz de Tejada et al., 2011	España (Canarias)	409	Coxartrosis	Media: 71,26 (DE ± 7,71)	63,3
HKQ ²⁰	Castellet et al., 2014	España (no especificado)	316	Coxartrosis con indicación de prótesis de rodilla antes y después de la cirugía	Media: 71,26 (DE ± 7,71)	–
HOS ²¹	Seijas et al., 2014	España (no especificado)	100	Pinzamiento acetabular mixto (37), pinzamiento tipo Cam (26), Cam +labrum (15), Tönnis II (5), Tönnis I (5), pinzamiento tipo Pincer (4), lesión del labrum (3), trocanteritis (2), secuelas de epifisiólisis (1), secuela de Perthes (1) y osteonecrosis (1)	Media: 45,1 (DE ± 12,1) Rango: 18-65	36
IHOT-33 ²²	Ruk-Ibán et al., 2015	España (no especificado)	97	Pinzamiento acetabular (78: 65 mixto, 11 tipo Cam, 2 tipo Pincer), artrosis (10), patología glúteo medio (3), secuelas de Perthes (2), secuelas de epifisiólisis (2), tendinitis del psoas (1), displasia congénita de cadera (1) y osteonecrosis (1)	Media: 43,8 (DE ± 10,9) Rango: 22-60	38,1

AMICAL: *Arthrose des Membres Inférieurs et Qualité de vie*; DE: desviación estándar; HKQ: *Hip and Knee Questionnaire*; HOS: *Hip Outcome Score*; IHOT-33: *International Hip Outcome Tool-33*; LEFS: *Lower Extremity Functional Scale*; LLFI: *Lower Limb Functional Index*; OAKHQOL: *Osteoarthritis Knee and Hip Quality of Life questionnaire*; WOMAC: *Western Ontario and McMaster Universities Osteoarthritis*.

Tabla 2
Evaluación de la metodología empleada para la adaptación transcultural de los cuestionarios

Escala	Traducción	Síntesis	Traducción inversa	Análisis por comité de expertos	Prueba piloto (número de pacientes)
LLFI ¹⁵	+	+	–	+	0
LEFS ¹⁶	+	+	+	+	0
AMICAL ¹⁷	+	+	+	+	+(10)
WOMAC ¹⁸	+	+	–	+	+(10)
OAKHQOL ¹⁹	+	+	+	+	+(17)
HKQ ²⁰	+	+	–	7	+(20)
HOS ²¹	+	+	+	+	+(30)
IHOT-33 ²²	+	+	+	+	+(30)

(+): correctamente hecho; (?): dudoso; (–): incorrectamente hecho o no hecho; (0): no se da información sobre si se hizo o no se hizo; AMICAL: *Arthrose des Membres Inférieurs et Qualité de vie*; HKQ: *Hip and Knee Questionnaire*; HOS: *Hip Outcome Score*; IHOT-33: *International Hip Outcome Tool-33*; LEFS: *Lower Extremity Functional Scale*; LLFI: *Lower Limb Functional Index*; OAKHQOL: *Osteoarthritis Knee and Hip Quality of Life questionnaire*; WOMAC: *Western Ontario and McMaster Universities Osteoarthritis*.

valores, para TE y RME, que indicaban una sensibilidad grande en ambos casos.

En el cuestionario AMICAL¹⁷ la adaptación transcultural se realizó en un país latinoamericano (México). No hay diferencias culturales tan marcadas que dificulten el uso del cuestionario en España. La introducción de pequeños cambios en algunas palabras usadas permitiría emplearlo. Por ejemplo, se podría sustituir la palabra «pesero» por «autobús».

Discusión

El objetivo ha sido revisar sistemáticamente los cuestionarios autocumplimentados para patologías de cadera adaptados al español analizando la calidad metodológica del proceso de adaptación transcultural y verificando las propiedades psicométricas de

la nueva versión obtenida. En la revisión hemos encontrado 8 cuestionarios adaptados a nuestro idioma¹⁵⁻²². En todos los casos el procedimiento de adaptación se realizó correctamente, siguiendo las recomendaciones internacionales³⁻⁵ y asegurando la equivalencia conceptual con la versión original desarrollada en otro idioma. El rigor fue máximo para 5 cuestionarios^{16,17,19,21,22}. Las propiedades psicométricas analizadas fueron aceptables y, en general, similares a las de la versión original y a las de otras versiones de la escala adaptadas a otros idiomas. Aunque en ningún estudio se valoraron todas las propiedades psicométricas posibles, la fiabilidad y la validez se analizaron, de diversos modos, en todos los cuestionarios.

En las últimas décadas se han desarrollado y validado numerosas escalas para patologías de cadera²⁴⁻³⁶. Es preferible adaptar un cuestionario ya existente, comprobando que la nueva versión mantiene las propiedades psicométricas de la original, que crear

Tabla 3
Principales propiedades psicométricas analizadas en los cuestionarios adaptados

Escala	Consistencia interna (α de Cronbach) total	Reproducibilidad test-retest: número de participantes (tiempo entre las 2 evaluaciones) ICC	Instrumento de comparación: validez (coeficiente de correlación)	Efectos suelo/techo	Sensibilidad	
					TE	RME
LLFI ¹⁵	0,91	— 0,96 (rango de 0,93 a 0,97)	WOMAC: fuerte ($r = 0,77$) EQ-5D-3L: justo e inversamente correlacionados ($r = -0,62$)	—	—	—
LEFS ¹⁶	0,98	— 0,998 (IC 95%: 0,996-0,999)	SF-36: funcionalidad ($r = 0,896$); rol físico ($r = 0,800$); rol emocional ($r = 0,504$); social ($r = 0,764$); salud mental ($r = 0,744$); vitalidad ($r = 0,760$); dolor ($r = 0,903$); salud general ($r = 0,657$); componente físico ($r = 0,902$); componente mental ($r = 0,824$); $p < 0,01$	No/No	2,3	1,9
AMICAL ¹⁷	0,946	65 (72 h) 0,979 (IC 95%: 0,961-0,991)	WOMAC: EVA ($r = 0,737$); capacidad funcional ($r = -0,831$); dolor ($r = -0,742$); rigidez ($r = -0,649$); $p < 0,05$ SF-36: actividad física ($r = 0,709$); rol físico ($r = 0,678$); salud mental ($r = 0,705$); salud general ($r = 0,506$); rol emocional ($r = 0,432$). En otros dominios, las correlaciones fueron bajas pero significativas: dolor ($r = 0,484$); función social ($r = 0,346$); apoyo social ($r = 0,287$); $p < 0,05$	—	—	—
WOMAC ¹⁸	Subescala: Dolor: 0,71 a 0,97 Función: 0,64 a 0,95	17 (72 h) Subescala: Dolor: 0,36 a 0,76 Función: 0,34 a 0,52	WOMAC: dolor ($r = 0,27$); función ($r = 0,77$)	—	—	—
OAKHQOL ¹⁹	Rango de 0,60 a 0,93	— Rango de 0,75 a 0,81	SF-36: moderadamente correlacionado con los dominios físicos (función física, $r = 0,58$; dolor, $r = 0,64$; salud general, $r = 0,41$; y PCS, $r = 0,59$) WOMAC: moderadamente correlacionado con la función ($r = -0,80$) y el dolor ($r = -0,69$) EQ-5D: $r = 0,5$	—	—	—
HKQ ²⁰	0,864	No estudiado	WOMAC: dolor y funcionalidad, $r = 0,641$; rigidez, $r = 0,533$	No/No	—	—
HOS ²¹	Subescala: AVD: 0,95 Deporte: 0,9	— (15 días) Subescala: AVD: 0,95 (IC 95%: 0,92-0,97) Deporte: 0,94 (IC 95%: 0,89-0,97)	WOMAC: Subescala AVD: dolor ($r = 0,699$); rigidez ($r = 0,667$); función ($r = 0,788$) Subescala deporte: dolor ($r = 0,607$); rigidez ($r = 0,499$); función ($r = 0,764$)	No/Si (para subescala AVD deporte)	Subescala: AVD — Deporte: —	1,53 — 1,27 1,18
iHOT-33 ²²	Global: 0,98 Subescala: Función: 0,97 Deporte: 0,94 Trabajo: 0,89 Social: 0,94	— (15 días) Global: 0,97 (IC 95%: 0,96-0,99) Subescala: Función: 0,95 (IC 95%: 0,92-0,98) Deporte: 0,92 (IC 95%: 0,76-0,98) Trabajo: 0,93 (IC 95%: 0,83-0,98) Social: 0,96 (IC 95%: 0,91-0,98)	WOMAC: Subescala funcional: dolor ($r = 0,744$); rigidez ($r = 0,687$); función ($r = 0,79$) Subescala deporte: dolor ($r = 0,526$); rigidez ($r = 0,513$); función ($r = 0,536$) Subescala trabajo: dolor ($r = 0,581$); rigidez ($r = 0,544$); función ($r = 0,616$) Subescala social: dolor ($r = 0,653$); rigidez ($r = 0,574$); función ($r = 0,658$)	Si/No	—	—

Los valores negativos de los coeficientes de correlación indican correlación inversa.
AMICAL: *Arthrose des Membres Inférieurs et Qualité de vie*; AVD: actividades de la vida diaria; EQ-5D-3L: *European Health Questionnaire 5 Dimensions 3 Levels*; EVA: escala visual analógica; HKQ: *Hip and Knee Questionnaire*; HOS: *Hip Outcome Score*; IC: intervalo de confianza; ICC: coeficiente de correlación intraclass; iHOT-33: *International Hip Outcome Tool-33*; LEFS: *Lower Extremity Functional Scale*; LLFI: *Lower Limb Functional Index*; OAKHQOL: *Osteoarthritis Knee and Hip Quality of Life questionnaire*; PCS: *Physical Component Summary*; RME: respuesta media estandarizada; SF-36: *Short Form 36 Health Survey*; TE: tamaño del efecto; WOMAC: *Western Ontario and McMaster Universities Osteoarthritis*.

una nueva. La adaptación es un proceso más económico y el uso de un mismo cuestionario facilita comparaciones entre diferentes poblaciones. Los 7 cuestionarios adaptados en España podrían emplearse sin problemas en países iberoamericanos. No obstante, en algunos casos será necesario algún pequeño cambio en alguna de las palabras usadas para adaptarse a la forma local de hablar en algunas naciones. Lo mismo sucede en el caso del cuestionario AMICAL adaptado en México¹⁷ para su uso en España o en otros países latinoamericanos. El hecho de que en la mayoría de los estudios los participantes fueran predominantemente mujeres^{15,17-20}, salvo en 3, no parece impedimento para utilizar los cuestionarios en varones.

Los 8 cuestionarios adaptados al español presentan características distintas y la elección de uno u otro variará según las necesidades y el contexto, ya sea este clínico y/o de investigación.

Para seleccionar qué cuestionario utilizar hay que considerar no solo los criterios metodológicos (propiedades psicométricas), sino también diversos aspectos prácticos como el tiempo necesario para cumplimentarlo y puntuarlo y la utilidad para determinadas patologías y/o determinados grupos de pacientes.

De los 8 cuestionarios identificados en la revisión había 3 para patologías de todo el miembro inferior (LLFI, LEFS y AMICAL)¹⁵⁻¹⁷, otros 3 para patologías de rodilla y/o cadera (WOMAC, OAKHQOL y HKQ)¹⁸⁻²⁰ y 2 exclusivamente para patología de cadera (HOS y iHOT-33)^{21,22}. Los más usados internacionalmente son LEFS y WOMAC.

El cuestionario LEFS incluye 20 preguntas. A diferencia del WOMAC, permite discriminar entre dolor y función física. Se cumplimenta en 2 min y se puntúa en 20 s. El cuestionario LLFI, de 25 ítems, todavía se ha empleado poco debido a su muy reciente

creación. El cuestionario AMICAL es el único específico para valorar la calidad de vida de los pacientes con artrosis de miembros inferiores. Su desventaja es que incluye 43 cuestiones y el paciente tarda en rellenarlo alrededor de 15 min.

El cuestionario WOMAC es el más usado internacionalmente para valorar la capacidad funcional de pacientes con artrosis de rodilla y/o cadera y para evaluar la eficacia de los resultados de tratamientos conservadores y quirúrgicos. Se compone de 24 ítems y el paciente tarda en contestarlo alrededor de 5-10 min. Se han elaborado 2 versiones reducidas, con solo 11 y 14 cuestiones, con el objetivo de conseguir la misma información que el cuestionario original pero reduciendo el tiempo de administración. Se ha comprobado que ambas versiones en español conservan las mismas propiedades psicométricas que la versión de 24 ítems³⁷⁻³⁹. Estas nuevas versiones pueden ser empleadas en la práctica clínica y en la investigación. Como aspectos negativos, la escala WOMAC no evalúa aspectos relacionados con la calidad de vida y no permite discriminar cuál es la articulación responsable en caso de que la cadera y la rodilla estén afectadas simultáneamente. Sus puntuaciones pueden verse modificadas por la presencia de dolor lumbar, alteraciones en el estado de ánimo y otras patologías extraarticulares⁴⁰. El cuestionario OAKHQOL es muy completo ya que incluye aspectos relacionados con salud mental, apoyo social, función social, relaciones personales, actividad sexual y vida profesional. El elevado número de ítems que lo componen dificulta su empleo en la práctica clínica. El cuestionario HKQ consta de 16 preguntas. Su validación en español se realizó solo en artroplastias de rodilla, sin incluir pacientes con patologías de cadera. Eso podría representar un inconveniente al aplicarlo en ese tipo de pacientes.

Los cuestionarios clásicos diseñados para valorar resultados en pacientes mayores con artrosis de cadera son menos útiles en individuos jóvenes y activos con patología no artrósica y con mayores demandas funcionales. Disponemos de 2 cuestionarios adaptados para estos pacientes. El cuestionario HOS contiene 28 ítems y evalúa específicamente aspectos funcionales. En un reciente metaanálisis³⁴ se recomienda su uso para valorar los resultados tras artroscopia de cadera. El cuestionario iHOT-33 valora, a diferencia del HOS, aspectos emocionales, sociales, laborales y del estilo de vida. Existe una versión reducida de 12 preguntas (iHOT-12)⁴¹ que podría ser más viable en la clínica aunque aún no está validada en español.

La revisión es especialmente relevante porque 6 de las 8 adaptaciones se han publicado en los 2 últimos años. Como fortalezas de nuestra revisión podríamos destacar, en primer lugar, que tanto el rigor del proceso de adaptación del cuestionario al español como las propiedades psicométricas analizadas fueron examinadas por 2 autores. Además, el proceso de búsqueda fue exhaustivo, por lo que resulta improbable que no se hayan encontrado otras escalas adaptadas al español que estén publicadas. Aunque se seleccionaron las bases de datos más utilizadas, es posible que algunos estudios no se hayan detectado, ya que algunas revistas no pueden ser indexadas en estas bases de datos.

En conclusión, actualmente disponemos de 8 escalas para medir la limitación funcional de la cadera que han sido adaptadas a la población española siguiendo las recomendaciones internacionales (en 5 casos con el máximo rigor). Todas tienen propiedades psicométricas aceptables. La mayor parte de las adaptaciones (6 de las 8) se han publicado en los 2 últimos años (a partir de 2014). Algunas fueron diseñadas para población adulta y otras para problemas típicos de pacientes más jóvenes (principalmente el pinzamiento femoroacetabular). Algunas se centran específicamente en la valoración funcional (más cortas y sencillas, pero válidas para la práctica clínica habitual), mientras que otras abarcan múltiples aspectos (emocionales, sociales, laborales, etc.), lo que las hace especialmente indicadas para

la evaluación de casos complejos o con objetivos de investigación.

Responsabilidades éticas

Protección de personas y animales. Los autores declaran que para esta investigación no se han realizado experimentos en seres humanos ni en animales.

Confidencialidad de los datos. Los autores declaran que han seguido los protocolos de su centro de trabajo sobre la publicación de datos de pacientes.

Derecho a la privacidad y consentimiento informado. Los autores declaran que en este artículo no aparecen datos de pacientes.

Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

Bibliografía

1. Badía X, Salameo M, Alonso J. La medida de la salud: guía de escalas de medición en español. 2.^a ed. Barcelona: Edimac; 1999.
2. Noble PC, Fuller-Latfeniere S, Meftah M, Dwyer MK. Challenges in outcome measurement: Discrepancies between patient and provider definitions of success. *Clin Orthop Relat Res.* 2013;471:3437-45.
3. Muñoz J, Elosua P, Hambleton RK. Directrices para la traducción y adaptación de los tests: segunda edición. *Psicothema.* 2013;25:151-7.
4. Epstein J, Santo RM, Guillemin F. A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus. *J Clin Epidemiol.* 2015;68:435-41.
5. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-related quality of life measures: Literature review and proposed guidelines. *J Clin Epidemiol.* 1993;46:1417-32.
6. Scholtes VA, Terwee CB, Postman RW. What makes a measurement instrument valid and reliable? *Injury.* 2011;42:236-40.
7. Instituto Cervantes. Informe: El español una lengua viva; 2015 [consultado 6 Feb 2016]. Disponible en: <http://www.cervantes.es/imagenes/File/prensa/EIX20espaolG20unaL20lenguaG20viva.pdf>
8. Murphy DR, Lopez M. Neck and back pain specific outcome assessment questionnaires in the Spanish language: A systematic literature review. *Spine J.* 2013;13:1667-74.
9. Altman DG. *Practical statistics for medical research.* London: Chapman & Hall; 1999.
10. 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. *J Clin Epidemiol.* 2007;60:34-42.
11. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976).* 2000;25:3186-91.
12. 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 Health.* 2005;8:94-104.
13. Ramada-Rodilla JM, Serra-Pujadas C, Delclos-Clanchet GL. Adaptación cultural y validación de cuestionarios de salud: revisión y recomendaciones metodológicas. *Salud Pública Mex.* 2013;55:57-66.
14. 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.* 2012;21:651-7.
15. Cuesta-Vargas AI, Gabel CP, Bennett P. Cross cultural adaptation and validation of a Spanish version of the Lower Limb Functional Index. *Health Qual Life Outcomes.* 2014;12:75.
16. Cruz-Díaz D, Lomas-Vega R, Osuna-Pérez MC, Hita-Contreras F, Fernández AD, Martínez-Amat A. The Spanish lower extremity functional scale: A reliable, valid and responsive questionnaire to assess musculoskeletal disorders in the lower extremity. *Disabil Rehabil.* 2014;36:2005-11.
17. Espinosa-Cuervo G, Guillemin F, Rat AC, Duarte-Salazar C, Alemán-Hernández SI, Vergara-Álvarez Y, et al. Transculturalization and validation of a Spanish translation of the specific lower limb osteoarthritis and quality of life questionnaire AMICAL: Arthrose des Membres Inférieurs et Qualité de vie AMIQUAL. *Reumatol Clin.* 2014;10:241-7.
18. Villanueva I, Guzman MM, Toyos FJ, Ariza-Ariza R, Navarro F. Relative efficiency and validity properties of a visual analogue vs a categorical scaled version of the Western Ontario and McMaster Universities Osteoarthritis (WOMAC) Index: Spanish versions. *Osteoarthritis Cartilage.* 2004;12:225-31.

19. González Sáenz de Tejada M, Escobar A, Herdman M, Herrera C, García L, Sarasqueta C. Adaptation and validation of the Osteoarthritis Knee and Hip Quality of Life (OKHQOL) questionnaire for use in patients with osteoarthritis in Spain. *Clin Rheumatol.* 2011;30:1563-75.
20. Castellet E, Ares O, Celaya F, Valentí-Azcárate A, Salvador A, Torres A, et al. SERIOD group. Transcultural adaptation and validation of the Hip and Knee questionnaire into Spanish. *Health Qual Life Outcomes.* 2014;12:76.
21. Seijas R, Sallent A, Ruiz-Ibán MA, Ares O, Marín-Peña O, Cuñillar R, et al. Validation of the Spanish version of the Hip Outcome Score: A multicenter study. *Health Qual Life Outcomes.* 2014;12:70.
22. Ruiz-Ibán MA, Seijas R, Sallent A, Ares O, Marín-Peña O, Muriel A, et al. The International Hip Outcome Tool-33 (iHOT-33): Multicenter validation and translation to Spanish. *Health Qual Life Outcomes.* 2015;13:62.
23. Batlle-Cuadra E, Esteve-Vives J, Pieta MC, Hargreaves R, Cuts J. Traducción y adaptación al español del cuestionario WOMAC específico para artrosis de rodilla y cadera. *Rev Esp Reumatol.* 1999;25:38-45.
24. Ahmad MA, Kyriotes FN, Giannoudis PV. Measuring hip outcomes: Common scales and checklists. *Injury.* 2011;42:259-64.
25. Nilsson A, Bremander A. Measures of hip function and symptoms: Harris Hip Score (HHS), Hip Disability and Osteoarthritis Outcome Score (HOOS), Oxford Hip Score (OHS), Lequesne Index of Severity for Osteoarthritis of the Hip (LSDH), and American Academy of Orthopedic Surgeons (AAOS) Hip and Knee Questionnaire. *Arthritis Care Res (Hoboken).* 2011;53 Suppl. 11:S200-7.
26. Bryant DM, Sanders DW, Coles CP, Petrisor BA, Jeray RJ, Lallamne GY. Selection of outcome measures for patients with hip fracture. *J Orthop Trauma.* 2009;23:434-41.
27. Ashby E, Crocetti MP, Haddad FS. Outcome measures for orthopaedic interventions on the hip. *J Bone Joint Surg Br.* 2008;90:545-9.
28. Harris K, Dawson J, Gibbons E, Lim CR, Beard DJ, Fitzpatrick R, et al. Systematic review of measurement properties of patient-reported outcome measures used in patients undergoing hip and knee arthroplasty. *Patient Relat Outcome Meas.* 2016;7:101-8.
29. Riddle DL, Stratford PW, Bowman DH. Findings of extensive variation in the types of outcome measures used in hip and knee replacement clinical trials: A systematic review. *Arthritis Rheum.* 2008;50:876-83.
30. Alviar MJ, Oliver J, Brand C, Tropea J, Hale T, Pirpiris M, et al. Do patient-reported outcome measures in hip and knee arthroplasty rehabilitation have robust measurement attributes? A systematic review. *J Rehabil Med.* 2011;43:572-83.
31. Ramisetty N, Kwon Y, Mohtadi N. Patient-reported outcome measures for hip preservation surgery—a systematic review of the literature. *J Hip Preserv Surg.* 2015;2:15-27.
32. Kemp JL, Collins NJ, Roos EM, Crossley KM. Psychometric properties of patient-reported outcome measures for hip arthroscopic surgery. *Am J Sports Med.* 2013;41:2065-73.
33. Harris-Hayes M, McDonough CM, Leung M, Lee CB, Callaghan JJ, Roos EM. Clinical outcomes assessment in clinical trials to assess treatment of femoroacetabular impingement: Use of patient-reported outcome measures. *J Am Acad Orthop Surg.* 2013; Suppl. 1:S39-46.
34. Thorborg K, Roos EM, Bartels EM, Petersen J, Hölmich P. Validity, reliability and responsiveness of patient-reported outcome questionnaires when assessing hip and groin disability: A systematic review. *Br J Sports Med.* 2010;44:1186-96.
35. Thorborg K, Tjissen M, Habets B, Bartels EM, Roos EM, Kemp J, et al. Patient-Reported Outcome (PRO) questionnaires for young to middle-aged adults with hip and groin disability: a systematic review of the clinimetric evidence. *Br J Sports Med.* 2015;49:812.
36. Lodhia P, Slobogean GP, Noonan VK, Gilhart MK. Patient-reported outcome instruments for femoroacetabular impingement and hip labral pathology: A systematic review of the clinimetric evidence. *Arthroscopy.* 2011;27:279-86.
37. López Alonso SR, Martínez Sánchez CM, Romero Cañadillas AB, Navarro Casado F, González Rojo J. Propiedades métricas del cuestionario WOMAC y de una versión reducida para medir la sintomatología y la discapacidad física. *Aten Primaria.* 2009;41:613-20.
38. Escobar A, Vrotsos K, Bilbao A, Quintana JM, García-Pérez I, Herrera-Espiteira C. Validación de una escala reducida de capacidad funcional del cuestionario WOMAC. *Gac Sanit.* 2011;25:513-8.
39. Bilbao A, Quintana JM, Escobar A, Las Hayas C, Orive M. Validation of a proposed WOMAC short form for patients with hip osteoarthritis. *Health Qual Life Outcomes.* 2011;9:75.
40. Sánchez-Sotelo J. Instrumentos de valoración del estado de la salud en Traumatología y Cirugía Ortopédica. *Rev Ortop Traumatol.* 2004;48:304-14.
41. Griffin DR, Parsons N, Mohtadi NG, Safran MR. Multicenter Arthroscopy of the Hip Outcomes Research Network. A short version of the International Hip Outcome Tool (iHOT-12) for use in routine clinical practice. *Arthroscopy.* 2012;28:611-6.

