

eman ta zabal zazu



Universidad  
del País Vasco

Euskal Herriko  
Unibertsitatea

# Escalas de calidad de vida relacionada con la salud y funcionalidad en pacientes con patologías de hombro: revisión sistemática, evaluación estandarizada y validación de un cuestionario

---

Kalliopi Vrotsou

Donostia, 2017

Tesis doctoral dirigida por el Dr. Antonio Escobar Martínez

y el Dr. Juan José Aurrekoetxea Agirre



Στον Joseba, την Κλειώ και τον Κώστα



## Agradecimientos

Quiero agradecer a mis dos directores de tesis, por toda la ayuda, apoyo y comentarios constructivos durante la realización de este proyecto. A Antonio por “prestarme”, hace ya varios años, la idea inicial de *Hombro*, que fue el motor de todo este trabajo. A Juanjo, por aceptar involucrarse en este proceso, por sus aportaciones, mejoras y positivismo.

También quiero agradecer a los traumatólogos que colaboraron en el reclutamiento de pacientes y recogida de datos, ya que sin ellos ninguno de los proyectos que se presentan en este manuscrito hubiera sido posible. En especial a Ricardo Cuellar, por compartir mi entusiasmo por el proyecto de la escala *Constant*, responder a mis innumerables dudas sobre las patologías de hombro, y supervisar las decisiones que requerían la visión de un especialista.

A mis excompañeros, de las Unidades de Investigación de los Hospitales Basurto, y Galdakao-Usansolo, por su cálida acogida, por todo lo que he aprendido con ellos, por solucionarme problemas y abrirme puertas en el camino de la investigación. Especialmente a Ziortza por gestionar todo lo relacionado con la recogida de datos.

A los miembros de BiblioPRO, por la ayuda con la administración de la herramienta EMPRO, y por participar en las evaluaciones de expertos y en la redacción de artículos.

A mis compañeras actuales de la Unidad de Investigación de AP-OSIS de Gipuzkoa, mil gracias por vuestro cariño, incansable apoyo, ayuda, y sobre todo por el trabajo bien hecho. A Bakartxo, Mónica y Mainer por participar activamente en diferentes fases de los estudios aquí presentados. A Itziar le quería agradecer además, su apoyo constante durante este largo proyecto y por permitir mi “ausencia” de la última temporada.

Agradecer también a todos los pacientes que participaron en nuestro estudio, por dedicar su tiempo para contestar a nuestras preguntas. Por último al Departamento de Sanidad del Gobierno Vasco, ya que sin su financiación, este proyecto no se hubiera realizado.

## Índice de Contenido

<b>RESUMEN</b> .....	<b>1</b>
<b>1. Introducción</b> .....	<b>7</b>
<b>2. Hipótesis y objetivos</b> .....	<b>15</b>
2.1 Hipótesis.....	15
2.2 Objetivos.....	16
<b>3. Calidad de vida relacionada con la salud y funcionalidad</b> .....	<b>17</b>
3.1 Calidad de vida relacionada con la salud .....	17
3.1.1 Escalas de calidad de vida relacionada con la salud .....	18
3.2 Capacidad funcional basada en la movilidad y la observación clínica .....	24
3.2.1 Pruebas funcionales de observación clínica .....	24
3.3 Atributos y propiedades psicométricas .....	27
3.3.1 Modelo conceptual y de medida .....	28
3.3.2 Interpretabilidad .....	28
3.3.3 Facilidad de uso .....	29
3.3.4 Formas alternativas de administración .....	30
3.3.5 Proceso de adaptación cultural & lingüística .....	30
3.3.6 Fiabilidad.....	30
3.3.7 Validez .....	31
3.3.8 Sensibilidad al cambio.....	33
3.4 Herramientas de evaluación de las escalas .....	34
3.4.1 GraQol.....	34
3.4.2 EMPRO.....	35
3.4.3 COSMIN.....	37
<b>4. Material y Métodos</b> .....	<b>39</b>
4.1 Estudio de evaluación de las escalas auto-cumplimentadas.....	40
4.2 Estudio de validación al castellano de la escala ASES-p.....	40
4.3 Estudio de evaluación de la escala Constant-Murley Score .....	43
4.4 Herramientas y escalas utilizadas en los diferentes estudios.....	44
4.5 Análisis estadístico .....	50

4.5.1	Fiabilidad.....	51
4.5.2	Validez.....	53
4.5.3	Sensibilidad al cambio.....	60
4.5.4	Software.....	61
<b>5.</b>	<b>Escalas auto-cumplimentadas para la patología de hombro: revisión sistemática y evaluación estandarizada de la evidencia disponible.....</b>	<b>63</b>
5.1	Introducción.....	64
5.2	Material y Métodos.....	65
5.3	Resultados.....	67
5.4	Discusión.....	77
<b>6.</b>	<b>Sección auto-cumplimentada del cuestionario ASES: validación al castellano aplicando métodos de la teoría clásica y el modelo Rasch.....</b>	<b>81</b>
6.1	Introducción.....	82
6.2	Metodología.....	82
6.3	Resultados.....	88
6.4	Discusión.....	97
<b>7.</b>	<b>Fiabilidad test-retest del cuestionario ASES-p.....</b>	<b>103</b>
7.1	Introducción.....	104
7.2	Metodología.....	104
7.3	Resultados.....	105
7.4	Discusión.....	108
<b>8.</b>	<b>Cualidades psicométricas de la escala Constant-Murley Score.....</b>	<b>111</b>
8.1	Introducción.....	112
8.2	Material y Métodos.....	113
8.3	Resultados.....	116
8.4	Discusión.....	120
<b>9.</b>	<b>Discusión general.....</b>	<b>125</b>
<b>10.</b>	<b>Conclusiones generales.....</b>	<b>141</b>
	<b>Bibliografía.....</b>	<b>143</b>
	<b>Anexos.....</b>	<b>159</b>





## RESUMEN

La articulación de hombro suele presentar una patología frecuente y variada. En nuestro ámbito la prevalencia de esta patología se sitúa entre 70-200 casos por 1000 adultos. Asimismo, se estima que los problemas relacionados con esta articulación, se encuentran entre los más frecuentes motivos de consulta en Atención Primaria. Los problemas de hombro están asociados a una serie de gastos médicos y económicos. Pero también, en los casos más severos, estas patologías pueden conllevar problemas en el ámbito laboral, social e incluso psicológico. Realizar un diagnóstico correcto es el primer paso para poder tratar adecuadamente cualquier problema de salud. Pero también, igual de importante, es disponer de herramientas capaces de “medir” el problema y proporcionar información valiosa sobre la efectividad de los tratamientos y la evolución de los pacientes.

El concepto de la “salud” y los indicadores del mismo han ido evolucionando a lo largo de los años. El aumento de la esperanza de vida y la aparición de nuevas patologías han hecho que la salud dejara de medirse en términos de mortalidad. En los años ‘40 la Organización Mundial de la Salud define la salud como *“un estado de completo bienestar físico, mental y social, y no solamente la ausencia de afecciones o enfermedades”*. Esta nueva visión impulsó el camino hacia el desarrollo de nuevas medidas en el ámbito de la

salud, las que hoy se conocen como escalas de Calidad de Vida Relacionada con la Salud (CVRS). La característica principal de estas escalas es su naturaleza subjetiva, ya que se basan en la percepción que los propios sujetos tienen sobre sus patologías. Las escalas de CVRS se clasifican en genéricas y específicas. Como su nombre indica, las primeras se pueden utilizar para valorar sujetos afectados de cualquier patología, mientras que las segundas se crean para valorar problemas de salud concretos. La naturaleza subjetiva de estos instrumentos, hace que no exista un consenso sobre la estructura, los componentes, el tipo y el número de preguntas (ítems) a incluir en los mismos. Hoy en día existen muchas escalas específicas de CVRS que evalúan diferentes patologías, tanto físicas como psíquicas. Pero, la falta de consenso arriba mencionada hace que además, para un mismo problema de salud, existan a menudo más de una escala.

Actualmente, en el ámbito de la articulación de hombro, nos encontramos con más de 20 instrumentos de CVRS. Algunas de ellos se crearon para valorar cualquier problema del miembro superior (mano/hombro/codo), otros han sido creados para patologías específicas, como la inestabilidad y otros se dirigen a pacientes con necesidades particulares, como los atletas o los sujetos en silla de ruedas. En relación a esta articulación existe un cuestionario que combina datos *subjetivos*, auto-cumplimentados por el paciente, con datos *objetivos*, basados en la observación clínica. Se trata de la escala funcional de Constant-Murley Score (CMS). El CMS se publicó en 1987 y, con la aprobación de la Sociedad Europea de Cirugía de Hombro y Codo (SECEC), se presentó como un instrumento adecuado para la valoración del hombro, independientemente de su patología. Este instrumento tiene una particular relevancia, ya que es uno de los más usados y referenciados en este ámbito. Asimismo, a lo largo de los años ha llegado a percibirse, por muchos, como “regla de oro” en la evaluación de esta articulación. Esto podría, por un lado, atribuirse al apoyo y promoción realizada por la SECEC. Mientras que, por otro lado, el “peso” que el componente de la observación clínica tiene en esta escala (i.e. 65% de su puntuación total), podría también explicar su amplia aceptabilidad por parte de los especialistas.

Para que una escala sea adecuada y recomendable, tiene que cumplir una serie de características, denominadas propiedades psicométricas. Estas propiedades constituyen

una parte fundamental en el estudio y valoración de los cuestionarios utilizados en el ámbito de la salud, ya que definen su calidad. La fiabilidad, la validez, y la sensibilidad al cambio, son las más conocidas y las que más atención reciben en el estudio de las escalas de medición. Pero existen atributos adicionales que también condicionan la calidad de los cuestionarios. El modelo conceptual y de medida, la interpretabilidad, la facilidad de uso, las formas alternativas de administración, y la adaptación cultural y lingüística, son características cuya presencia avala la utilidad de los instrumentos e indica los más apropiados entre ellos.

La presente tesis se desarrolla alrededor de la medición de la patología de hombro. Concretamente se han estudiado las características métricas de varias escalas multi-ítem de CVRS, y de la escala CMS por separado. También, como parte de este trabajo, se ha realizado la adaptación y validación castellana de un cuestionario auto-cumplimentado para la medición de la CVRS de las personas con patologías de hombro. Los resultados de los diferentes estudios aquí incluidos se han publicado en revistas internacionales: a) las características de las medidas de CVRS en el *Journal of Shoulder and Elbow Surgery*; b) la validación castellana del instrumento de CVRS en *Health and Quality of Life Outcomes*; c) la evaluación de las características de CMS está actualmente (mayo 2017) bajo evaluación, tras cambios mayores en la revista *Health Services Research*. Los estudios de esta tesis se realizaron con financiación conseguida por dos proyectos de investigación, del Departamento de Sanidad del Gobierno Vasco (2010111156, 201311086).

Cuando se inició el trabajo actual (diciembre 2010) no existía en nuestro entorno una escala específica para la patología de hombro que evaluara la CVRS de estos pacientes, mientras que el ámbito internacional contaba con un gran número de escalas auto-cumplimentadas para tal fin. De interés fueron aquellas escalas multi-ítem, que se pudieran utilizar para la valoración de hombro, independientemente de su patología, pero que fuesen específicamente creadas para esta articulación. Se excluyeron las escalas creadas para patologías concretas, y los cuestionarios que valoraban los problemas de hombro en pacientes con necesidades particulares. Asimismo, se excluyeron las escalas que evaluaban todo el miembro superior. Por este motivo, no se

consideró en nuestro estudio la adaptación castellana de *Disability of the Arm Shoulder and Hand (DASH)*. Para poder seleccionar el instrumento más adecuado, entre todos los existentes en la literatura, se realizó una búsqueda sistemática y posteriormente una evaluación estandarizada de la evidencia disponible. La búsqueda bibliográfica se realizó en PubMed y se consideraron solo artículos escritos en inglés. Se evaluaron 112 artículos en los que se encontró información sobre 11 escalas auto-cumplimentadas, específicamente creadas para la patología de hombro. La evaluación de la evidencia se basó en la herramienta *Evaluating the Measurement of Patient-Reported Outcomes (EMPRO)*, que valora las características métricas anteriormente mencionadas. Esta herramienta, realiza una evaluación por atributo y una global para cada escala, y asigna una puntuación por atributo y por escala. Dichas puntuaciones oscilan entre un mínimo de 0 a un máximo de 100 puntos; valores  $\geq 50$  puntos se consideran aceptables. Las evaluaciones se realizaron por pares de evaluadores, expertos en medidas de calidad de vida y todas fueron consensuadas. En relación a los diferentes atributos evaluados, de manera global, la sensibilidad al cambio, que es la capacidad de un instrumento en captar cambios en el tiempo, tras una intervención por ejemplo, fue la mejor evaluada. Por otro lado, la interpretabilidad, la capacidad de asignar un significado fácilmente comprensible a las puntuaciones que proporciona una escala, recibió la peor evaluación.

De las 11 escalas valoradas en este trabajo, 6 superaron la mínima puntuación aceptable. Entre ellas, 3 podrían ser buenos candidatos a considerar según el objetivo y las particularidades de los estudios. Las tres escalas mejor valoradas, fueron en este orden: *American Shoulder and Elbow Surgeons shoulder Assessment*, *Patient self-evaluation section (ASES-p)*, *Simple Shoulder Test (SST)* y *Oxford Shoulder Score (OSS)*. El ASES-p fue la finalmente seleccionada para ser adaptada y validada al castellano. Este instrumento cuenta con 11 ítems y dos sub-escalas (*dolor y función*). Además de haber alcanzado la mayor puntuación en nuestro estudio, este cuestionario es uno de los más utilizados en la investigación de la articulación de hombro. Y dicha la escala se encontraba ya adaptada y validada en varios otros idiomas. Se consideró que validar este instrumento al idioma diana, nos permitiría más flexibilidad a la hora de realizar comparaciones con datos obtenidos en otros ámbitos. Además se tomó en consideración el hecho de que la escala contaba con respuestas Likert, que suelen tener

mejor funcionamiento en este tipo de cuestionarios y permiten más flexibilidad en cuanto el análisis de datos, en comparación con las respuestas binarias que tenía el SST, por ejemplo. Finalmente, el hecho que esta escala evaluara el dolor como una dimensión con su propia puntuación, fue también de interés, ya que en la literatura se subraya la importancia que el elemento del dolor tiene en la patología de hombro.

Para el estudio de la validación de la escala ASES-p se reclutaron n=180 sujetos, mayores de 18 años, en las consultas de traumatología de cinco hospitales participantes. Los reclutados se sometieron a un tratamiento, quirúrgico o conservador, en el hombro afectado y se siguieron durante 6 y 3 meses tras finalizar el mismo, respectivamente. Se recibieron n=161 respuestas basales validas, de las cuales n=106 habían cumplimentado todos los ítems de la escala de interés. La consistencia interna de ASES-p fue adecuada, con alpha de Cronbach=0.91. La validez convergente y discriminante con las escalas SF-36 y CMS, presentó correlaciones >0.50. Los modelos de análisis factorial confirmatorio (AFC) y el modelo Rasch apoyaron la presencia de un factor latente. A pesar de ello, uno de sus ítems (deporte habitual) no fue informativo, aunque los resultados no indicaron su exclusión. La sensibilidad al cambio se evaluó mediante la estimación de efectos estandarizados, que indicaron que la puntuación total y de la sub-escala de la función son sensibles al cambio, aunque la sub-escala del dolor no presentó el mismo comportamiento. La reproducibilidad de la escala (fiabilidad test-retest) se estudió a posteriori, mediante entrevistas telefónicas realizadas a los participantes. Según estos resultados, aunque el ASES-p es reproducible, el *dolor* presenta límites de concordancia más anchos que la *función*. Los resultados no justifican la distribución de la puntuación total en las dos sub-escalas. También dirigen la atención hacia el ítem del deporte.

En relación al CMS y sus propiedades métricas, en este caso nuestra hipótesis previa, según lo publicado en la literatura, fue que estas propiedades serían distintas, en las diferentes patologías de la articulación. El CMS valora 4 componentes: *dolor, actividad de la vida diaria, movilidad y fuerza muscular*. Se realizaron búsquedas sistemáticas en PubMed y Embase y la evidencia se evaluó de nuevo por pares de expertos con la herramienta EMPRO. Tras revisar los 96 artículos finalmente incluidos, se consideraron 7 grupos: *patología subacromial, fracturas, artritis, inestabilidad, hombro congelado,*

*varias patologías y sujetos sanos.* De manera global, la sensibilidad al cambio fue la cualidad mejor valorada. La falta de información sobre consistencia interna es un dato a destacar, mientras que la interpretabilidad obtuvo en 3 ocasiones valores más elevados que la validez. Por último, no se ha podido encontrar justificación sobre la distribución de la puntuación total en los 4 componentes de la escala. Los datos disponibles apoyarían el uso de CMS solo en sujetos con patología subacromial, pero la evidencia evaluada no da soporte al amplio uso de este instrumento en todas las patologías de hombro.

Nuestros estudios tienen ciertas limitaciones. Las evaluaciones EMPRO pueden haberse condicionado por la experiencia de los evaluadores. La involucración de investigadores expertos en medidas de calidad de vida, la evaluación por pares y el consenso final deberían haber reducido este sesgo. Por otro lado, publicaciones no incluidas en la fase de la evaluación, también pueden haber afectado los resultados obtenidos. Las búsquedas manuales realizadas, deberían haber disminuido también este sesgo. Cabe mencionar que el número de sujetos, reclutados en la validación de ASES-p, no ha permitido realizar análisis para estudiar el comportamiento diferencial de los ítems de la escala, en diferentes grupos de pacientes (*differential item functioning*). Sería interesante estudiar posibles diferencias en cuanto al sexo, la edad o gravedad de la patología. Desgraciadamente este tipo de análisis requeriría de una muestra mucho mayor, actualmente no accesible considerando los recursos disponibles.

Se espera que los resultados de esta tesis resulten de utilidad para los clínicos e investigadores. La escala ASES-p podría utilizarse como instrumento de evaluación y de seguimiento de los pacientes con patologías de hombro, en nuestro entorno, aunque no se recomendaría el uso de la sub-escala del dolor como puntuación separada. Estudiar en más detalle la asignación de puntuaciones de esta escala, desarrollar variaciones para sujetos con diferentes niveles de actividad física y comprobar su funcionamiento en diferentes patologías podrían ser futuras líneas de investigación. En relación a la escala funcional CMS, sería deseable disponer de más datos, antes de usarla como regla de oro en la valoración de hombro. La estructura de la escala, su dimensionalidad y propiedades requerirían de más estudio.

## **1. Introducción**

El hombro es una de las articulaciones más complejas del ser humano. Cuenta con más flexibilidad que cualquier otra parte del cuerpo, pero esta ventaja lo hace también susceptible a varios tipos de lesiones, por lo que esta articulación presenta una patología frecuente y variada, que se asocia a importantes gastos médicos, económicos y sociales (Largacha et al. 2006, van der Heijden 1999, Michener & Leggin 2001, Virta et al. 2012), e incluso psicológicos (Paananen et al. 2011).

La mayoría de los pacientes que solicitan consulta por un problema de hombro, lo hacen por dolor o incapacidad funcional, ya que las patologías más comunes son traumáticas e inflamatorias o degenerativas. La traumática suele ser más frecuente en deportistas, pero las patologías inflamatorias o degenerativas son cada día más habituales, sobre todo a causa del envejecimiento de la población (Luime et al. 2004). Entre las patologías más frecuentes se encuentra el denominado hombro doloroso, cuya causa principal es la tendinitis y la rotura del manguito de los rotadores. La inestabilidad, el hombro congelado y la necrosis avascular se presentan con menos frecuencia. Las patologías de hombro pueden disminuir la funcionalidad de las personas afectadas, limitar las actividades de su vida diaria y de ocio y dificultar o incluso, en los casos más severos, impedir la actividad laboral y el autocuidado. Una consecuencia inevitable de todas

estas limitaciones es el empeoramiento de la calidad de vida de los pacientes, así como la creación de diferentes grados de dependencia (Beaton & Richards 1996, Marin-Gomez et al. 2006, Gartsman et al. 1998). Dependiendo de su naturaleza y de su gravedad, las patologías de hombro se pueden tratar con medicamentos analgésicos y antiinflamatorios, con fisioterapia, con infiltraciones, o con intervenciones quirúrgicas. Estos tratamientos se aplican sobre todo a pacientes con procesos degenerativos, como la artrosis, inflamatorios, como la artritis reumatoide, o traumatológicos como la rotura del manguito rotador (Castellote Olivito JM et al. 2000).

La prevalencia e incidencia de las patologías de hombro varía mucho, entre otros factores, según el origen de la información, el tipo del estudio, las patologías consideradas, el horizonte temporal, así como las características de la muestra, entre otros. Una revisión sistemática de 19 estudios procedentes de varios países sitúa el rango de las prevalencias puntuales estimadas del hombro doloroso entre 69-260 casos por 1000 adultos (Luime et al. 2004). En la misma revisión la estimación proporcionada por datos españoles es de 78 casos por 1000, con un intervalo de confianza del 95% que oscila entre 61-95 casos. Otros autores de nuestro ámbito presentan estimaciones de prevalencia entre 70-200 casos por 1000 adultos (Vicente-Herrero et al. 2009).

En los últimos 30 años, el conocimiento sobre la patología de hombro y especialmente en relación a las intervenciones quirúrgicas, ha avanzado mucho. Los avances tecnológicos han permitido la mejoría de las intervenciones y han contribuido al desarrollo de nuevas técnicas como la cirugía artroscópica. La artroscopia requiere tamaños de incisiones más pequeños, ha disminuido el tiempo operatorio y las comorbilidades post operatorias, y finalmente ha contribuido a la disminución de las estancias hospitalarias (Lane et al. 2014, Wang et al. 2015, Walton & Murrell 2012). Aunque optar por una artroscopia pueda depender de las características de la lesión o la patología del paciente, las ventajas proporcionadas hacen que la cirugía de hombro sea cada vez más habitual. Asimismo, la sustitución de la articulación del hombro es la intervención más común en pacientes con artrosis, después de la sustitución de rodilla y cadera, tanto en Reino Unido (Arthritis Research UK 2016), como en Estados Unidos (American Academy of Orthopaedic Surgeons 2013).



Lo arriba mencionado subraya la relevancia y el gran valor que puede tener para los clínicos e investigadores, el acceso a datos que aporten información sobre el estado de salud de los pacientes, los cambios a lo largo del tiempo y la efectividad de los tratamientos que se aplican a las patologías de hombro. La información relacionada con el estado de esta articulación, se puede obtener tanto por datos proporcionados por el propio paciente, vía cuestionarios auto-administrados, o por datos que se basan en observaciones clínicas.

Las escalas que recogen la información del paciente ponen el foco en los aspectos de la vida diaria de los individuos y en las limitaciones que supone padecer una patología, desde la perspectiva del propio paciente. Dichas escalas se denominan escalas de calidad de vida relacionada con la salud (CVRS), y se clasifican entre genéricas y específicas. Las genéricas, preguntan sobre el estado de salud en general y se pueden administrar a sujetos de cualquier patología, mientras que las específicas están diseñadas para valorar la calidad de vida de grupos de interés concretos. Las preguntas, (ítems) de estas escalas, se centran en la perspectiva de los encuestados y la percepción que ellos tienen sobre los problemas, las limitaciones, y su estado actual (Mc Dowell 2006).

Al inicio de este proyecto, en diciembre de 2010, estudios anteriores realizados en nuestro entorno habían traducido y validado en castellano cuestionarios específicos para ciertas patologías osteomusculares. Entre ellos, el WOMAC, para pacientes que se someten a sustitución de cadera o rodilla (Escobar et al. 2002), el Neck Disability Index, para discapacidad cervical (Andrade Ortega et al. 2008) y la escala Roland-Morris para pacientes con problemas de espalda (Kovacs et al. 2002). También existían dos validaciones, en lengua castellana, del cuestionario *Disability of the Arm Shoulder and Hand* (DASH), realizadas por dos grupos de investigación. Una de ellas fue realizada con pacientes con patologías de la mano (Rosales et al. 2002) y la otra con pacientes que presentaban patologías de hombro, codo y mano conjuntamente (Hervas et al. 2006). Además una reciente publicación presentaba la adaptación castellana del *Wheelchair Users Shoulder Pain Index* (WUSPI), una escala específica para usuarios de silla de ruedas (Arroyo-Aljaro & Gonzalez-Viejo 2009). En aquel momento, en los estudios sobre la

patología del hombro, publicados por autores españoles, la valoración de los participantes se realizaba sobre todo mediante la escala funcional Constant-Murley Score (Constant & Murley 1987). Algunos autores también presentaban datos de cuestionarios genéricos, aplicables a cualquier patología, como el SF-36 (Torrens et al. 2000) o el EuroQol (Ginés et al. 2007), mientras que otros habían utilizado escalas no previamente validada al idioma diana (Guerra de Hoyos et al. 2004). No existía por lo tanto un cuestionario específico que evaluase la CVRS de los pacientes de hombro en nuestro ámbito, al mismo tiempo que en los estudios internacionales se implementaban varias escalas auto-administradas y específicas para dicho fin (Oh et al. 2009, Paul et al. 2004, Placzek et al. 2004, Angst et al. 2004).

La mayoría de los cuestionarios de CVRS para la patología de hombro fueron desarrollados en Reino Unido y Estados Unidos y estaban ya traducidos y validados en Alemán, Francés y Holandés entre otros idiomas (Castellote Olivito JM et al. 2000, Kirkley et al. 2003, Angst et al. 2004). Evaluar la enfermedad desde la perspectiva del paciente es una valiosa ayuda para la comprensión del problema, mientras que disponer de herramientas que se utilizan en otros países, permite también la realización de comparaciones internacionales, sobre un mismo problema de salud (McDowell & Jenkinson 1996, Alonso 2000).

Desde la perspectiva clínica, los datos funcionales relacionados con la patología de hombro, tienen que ver con aspectos sobre todo kinesiológicos, como pueden ser la rotación interna, externa y el rango de movimiento libre de dolor, o la fuerza de la propia articulación. Este tipo de información permite a los profesionales evaluar la capacidad funcional y la evolución de la articulación afectada. La recogida de la misma, requiere la participación de especialistas traumatólogos; hecho que hace que estos datos se perciban, por muchos, como datos *objetivos* (Hegedus et al. 2008).

El anteriormente mencionado *Constant Murley Score* (CMS), es uno de los instrumentos más relevantes en el ámbito de la investigación del hombro. El CMS se presentó en 1987 como una escala válida y aplicable a todas las patologías y condiciones de esta articulación (Constant & Murley 1987). Fue, desde su publicación, altamente

recomendado por la Sociedad Europea de Cirugía de Hombro y Codo (SECEC) (Barra-Lopez 2007) y a lo largo de los años se ha implementado en un gran número de estudios, siendo actualmente el cuestionario funcional más utilizado y referenciado en la literatura (Kirkley et al. 2003, Angst et al. 2011, Roe et al. 2013, Roy et al. 2010). El 65% de la puntuación total de CMS se basa en las observaciones del especialista, y el resto 35% en preguntas auto-cumplimentadas por el paciente (Constant et al. 2008). A pesar de su extenso uso, dicha escala cuenta con limitaciones comentadas por varios autores. Más en concreto, se desconocen los criterios implementados para la selección de sus ítems, la justificación sobre los puntos que cada ítem atribuye a la puntuación total, así como su fiabilidad y validez (Barra-Lopez 2007, Kirkley et al. 2003, Conboy et al. 1996). Asimismo, existe cierto escepticismo sobre la aplicabilidad de este cuestionario a todas las patologías de hombro (Conboy et al. 1996, Othman & Taylor 2004, van den Ende et al. 1996), o sin tener en consideración la edad y el sexo de los pacientes (Walton et al. 2007, Yian et al. 2005). Finalmente, también se ha criticado la falta de un proceso estandarizado a la hora de medir la fuerza de la articulación (Lillkrona 2008, Bankes et al. 1998). En un segundo artículo, el autor original de CMS aceptó que la escala no era apta para la evaluación del dolor agudo intermitente, que se suele observar en pacientes con luxación de hombro. Se propuso una puntuación ajustada por edad y sexo, y se ofrecieron instrucciones específicas sobre cómo evaluar el ítem de la fuerza. En el mismo artículo, el autor propuso además cambiar las alternativas de respuesta de ciertos ítems, pasando de respuestas categóricas a continuas (mediante una escala visual analógica) (Constant et al. 2008). Aunque el cuestionario CMS y sus respectivas propiedades han sido el objetivo de varias revisiones y evaluaciones, ningún autor ha presentado una evaluación estandarizada de las propiedades de este instrumento en las diferentes patologías de hombro (Kirkley et al. 2003, Angst et al. 2011, Roy et al. 2010, Slobogean & Slobogean 2011, Huang et al. 2015).

El desarrollo del presente trabajo se basó en la intención de abordar las carencias arriba mencionadas y, sobre todo, en aportar datos objetivos acerca de los instrumentos de evaluación de la articulación de hombro. Por un lado, se pretendía seleccionar el cuestionario de CVRS más apropiado para la patología del hombro, entre todos los existentes, para su posterior adaptación y validación al castellano. En esta fase nos

centramos en los cuestionarios creados para la valoración de la articulación del hombro en general. No se consideraron escalas creadas para patologías concretas del hombro, como son el Oxford Shoulder Instability (Dawson et al. 1999), o el Western Ontario Osteoarthritis of the Shoulder Index (Lo et al. 2001). Tampoco se consideraron escalas para todo el miembro superior, o para grupos concretos, como las anteriormente mencionadas DASH y WUSPI. El gran número de instrumentos existentes en la literatura (Kirkley et al. 2003, Angst et al. 2004), generó el problema de cómo elegir el más adecuado entre todos. Por este motivo se optó por realizar una revisión sistemática, y evaluar la evidencia disponible mediante la herramienta Evaluating Measures of Patient Reported Outcomes (EMPRO). Se trata de una herramienta estandarizada, creada para facilitar la evaluación de las escalas auto-cumplimentadas, en base a una serie de propiedades establecidas que tienen que cumplir estos instrumentos (Valderas et al. 2008). Las evaluaciones obtenidas indicaron el cuestionario ASES-p (Richards et al. 1994), como el mejor valorado entre todos los considerados (Schmidt et al. 2014). Tras estudiar los resultados obtenidos y considerar, entre otros, la relevancia y amplio uso del ASES-p, en el estudio de la patología del hombro (Longo et al. 2011, Kirkley et al. 2003, Slobogean & Slobogean 2011), se decidió proceder a su adaptación y validación al castellano.

Se pretendía, además, realizar de una evaluación estandarizada de las características de la escala CMS. En este caso también se realizó una revisión sistemática de la literatura, y la evidencia disponible se volvió a evaluar con la misma herramienta ante mencionada. A pesar del hecho de que el EMPRO fue creado para escalas auto-cumplimentadas y su aplicación para la evaluación de CMS podría percibirse como “inadecuada”, lo cierto es que en la literatura no existen herramientas especialmente creadas para la evaluación de escalas basadas en la observación clínica o de escalas que combinan ambos datos objetivos y subjetivos, como el CMS. Asimismo, los autores de un trabajo anterior que evaluó el funcionamiento de escalas y medidas de actividad física de ambos tipos, utilizaron una herramienta parecida al EMPRO, el COSMIN, para llevar a cabo su evaluación (Mitchell et al. 2013). En el estudio de CMS, presentado en esta tesis, los datos obtenidos no se evaluaron de manera global, sino en grupos de diferentes patologías del hombro. Este desglose, se realizó por dos motivos. Primero para

comprobar si el funcionamiento de la escala podría verse condicionado por la patología que padecen los pacientes que, tal y como ha sido señalado por otros autores. En segundo lugar, para que los resultados de este trabajo fuesen lo más comparables posible a un artículo anterior cuyo objetivo fue precisamente el estudio de las propiedades de CMS, aunque en aquel trabajo la síntesis de la evidencia se realizó de modo narrativo (Roy et al. 2010).

Información detallada sobre EMPRO se presenta en los capítulos 3 y 4 de la presente tesis. Los resultados relacionados con la evaluación, adaptación y validación de ASES-p al castellano se encuentran en los capítulos 5, 6 y 7. Los resultados de la revisión sistemática y evaluación de la escala CMS se encuentran en el capítulo 8.



## **2. Hipótesis y objetivos**

### **2.1 Hipótesis**

Mediante revisión sistemática y evaluación de la evidencia, con una herramienta estandarizada, se puede obtener información objetiva sobre las escalas de CVRS y de funcionalidad utilizadas en sujetos con patologías de hombro.

La evaluación realizada ayudaría a seleccionar la escala de CVRS más adecuada, para su posterior adaptación y validación al castellano. La adaptación cultural y la aplicación de diferentes metodologías estadísticas, apoyaría la calidad de las propiedades psicométricas de la escala finalmente seleccionada.

#### ***Hipótesis específicas***

1. Mediante una revisión sistemática y evaluación estandarizada, con la herramienta EMPRO, se pueden conocer las propiedades psicométricas de los cuestionarios multi-ítem, de CVRS para la patología de hombro.

2. La aplicación de diferentes métodos estadísticos, entre ellos el análisis factorial confirmatorio y del modelo Rasch, apoyarían la estructura subyacente, la validez, la fiabilidad y la sensibilidad al cambio de la versión castellana del cuestionario seleccionado.
3. Mediante una revisión sistemática y evaluación de la evidencia, con la herramienta EMPRO, se establecerían las características psicométricas de la escala funcional CMS en diferentes patologías de la articulación del hombro.

## **2.2 Objetivos**

Disponer de una herramienta válida y reproducible para valorar la CVRS de los sujetos con patologías de hombro y conocer las cualidades psicométricas de la escala funcional que se utiliza para la evaluación clínica de la misma articulación.

### ***Objetivos específicos***

1. Proporcionar a los clínicos e investigadores datos sobre las cualidades psicométricas de las diferentes escalas multi-ítem de CVRS para la patología de hombro.
2. Seleccionar, adaptar y validar al castellano la escala de CVRS más adecuada para la articulación de hombro.
3. Establecer las propiedades psicométricas de la escala funcional CMS, para diferentes patologías de la articulación del miembro superior.



### **3. Calidad de vida relacionada con la salud y funcionalidad**

#### **3.1 Calidad de vida relacionada con la salud**

Los primeros indicadores de salud de la población eran objetivos y medibles. En este contexto, la tasa de la mortalidad y especialmente la mortalidad infantil, fueron medidas muy utilizadas en la sociedad pre-industrial. Los cambios en el estilo de la vida que supuso la era industrial contribuyeron principalmente al aumento de la esperanza de vida. La mortalidad precoz y las enfermedades comunitarias dejaron de ser problemas prevalentes y se sustituyeron por la aparición de nuevas patologías, el aumento de la incapacidad y las enfermedades crónicas, sobre todo en los grupos poblacionales de edad más avanzada. Estos cambios hicieron que la salud dejara de medirse solamente en términos de mortalidad y que pasara a percibirse como *“el estado libre de enfermedad”* (Mc Dowell 2006, Rosser 1993).

En el año 1941 Sigerist opina que: *“La salud, no es simplemente la ausencia de la enfermedad: es algo positivo, una actitud alegre en la vida”*. Pocos años más tarde, en 1948 la Organización Mundial de la Salud (OMS) presenta una definición más amplia, con matices adicionales, incluyendo un componente social: *“La salud es un estado de*

*completo bienestar físico, mental y social, y no solamente la ausencia de afecciones o enfermedades".* Dichas definiciones subrayaban que los indicadores de salud puramente clínicos (pruebas de sangre, exploraciones físicas, imágenes digitalizadas etc.) no eran la única manera de estudiar la salud de las personas. Aunque inicialmente la definición de la OMS fue criticada por no medible, finalmente consiguió impulsar la investigación sobre nuevos indicadores de salud y fue la base para los conceptos de Calidad de Vida (CV) y Calidad de Vida Relacionada con la Salud (CVRS) (Mc Dowell 2006, Breslow 2006).

Con CV nos referimos a la percepción del individuo sobre las circunstancias de su vida, en el contexto de la sociedad que vive y los valores de la misma y siempre en relación con sus preocupaciones, estándares de vida, expectativas y retos finalmente cumplidos (The WHOQOL Group 1994). Por definición la CV es un concepto completamente subjetivo, que depende del propio individuo y la percepción que él tiene sobre los parámetros anteriormente mencionados (Mc Dowell 2006).

Asimismo, la CVRS es un concepto multidimensional que añade al concepto de CV, el estado funcional y psicológico, la capacidad de realizar tareas diarias y la interacción social entre otros (Guyatt et al. 1993). Esta característica hace que no exista una única definición consensuada del término CVRS, aunque sí existe un amplio acuerdo que dicho concepto se refiere a las consecuencias que un problema de salud y/o el tratamiento puede tener en la vida diaria de los individuos (Barofsky 2012, McDowell & Jenkinson 1996).

### **3.1.1 Escalas de calidad de vida relacionada con la salud**

Las escalas de CVRS fueron desarrolladas considerando que el paciente es capaz de proporcionar información válida y fiable sobre su estado de salud, así como sobre los beneficios de los tratamientos recibidos (Alonso 2000, McDowell & Jenkinson 1996). Mediante una serie de preguntas relevantes, se extrae información acerca del estado de la persona, sobre todo de aquellos aspectos de su vida diaria que se pueden ver afectados por la patología de interés. Los cuestionarios, pueden ser auto cumplimentados por el propio paciente, o con la participación de un entrevistador. La

auto valoración que realizan los individuos depende obviamente de muchos factores, no solo de los cambios de salud experimentados, sino también de aspectos como las expectativas previas al tratamiento, la satisfacción con las intervenciones y asistencia recibidas o su estilo de vida y las limitaciones que supone padecer una patología (Mc Dowell 2006).

Aunque la medición de la CVRS es una ciencia reciente, la capacidad de las escalas de CVRS a la hora de predecir eventos tan importantes como pueden ser la hospitalización, la muerte o la discapacidad se han descrito múltiples veces (Alonso 2000, Valderas et al. 2008). El hecho de que varias escalas pueden existir para una única patología, en combinación con el gran número de patologías físicas y psíquicas prevalentes, hace que el número de los instrumentos de CVRS existente en la literatura sea también muy elevado. Según una revisión sistemática, en el año 2002 existían por lo menos 1.275 instrumentos de este tipo (Garratt et al. 2002). Estos datos subrayan el gran valor y la relevancia que tiene para los clínicos e investigadores el concepto y la medición de la CVRS. Esta medición de la CVRS se realiza principalmente por dos tipos de escalas, las genéricas y las específicas.

### ***Escalas genéricas***

Las escalas genéricas miden la capacidad en realizar las actividades habituales, el estado de salud mental y/o las relaciones sociales de los encuestados. Los conceptos que estos instrumentos evalúan no son específicos de ninguna patología, grupo poblacional o sexo, sino genéricos que se pueden utilizar para valorar el estado de salud global de cualquier individuo. Este tipo de cuestionarios se suele implementar en estudios epidemiológicos y encuestas poblacionales. Permite comparar la calidad de vida de sujetos sanos o con diferentes patologías, mientras que también se utilizan como medidas de comparación en validaciones de cuestionarios para patologías concretas. Su uso es ampliamente reconocido y referenciado en la literatura científica (Cid-Ruzafa & Damian-Moreno 1997, Brazier et al. 1992, Alonso et al. 1998, Mc Dowell 2006, Ware, Jr. et al. 1998). Entre los cuestionarios genéricos más utilizados se encuentran el 36-item Short-Form Health Survey (SF-36) (Ware, Jr. & Sherbourne 1992), el Barthel (Mahoney &

Barthel 1965), el WHOQOL (The WHOQOL Group 1993) y el EuroQol (The EuroQol Group 1990).

El SF-36 recoge información sobre aspectos positivos y negativos de la salud y proporciona un perfil del estado de la misma. Evalúa diferentes dimensiones, frecuentemente presentes en los cuestionarios de este ámbito; entre ellas, la función física, la vitalidad, el dolor corporal y la salud mental. La amplia evidencia disponible, en relación a este cuestionario, apoya su uso en el ámbito de la investigación y la práctica clínica habitual (Vilagut et al. 2005). El índice Barthel valora la capacidad de llevar a cabo 10 actividades básicas de la vida diaria (ABVD). Sus ítems preguntan, entre otros, sobre el aseo personal, vestirse, desplazarse, subir y bajar escaleras. En nuestro ámbito, esta escala se utiliza tanto en atención primaria, como en los servicios de rehabilitación y de geriatría (Osakidetza Servicio Vasco de Salud 2017). El cuestionario WHOQOL fue el resultado de una colaboración internacional, apoyada por la OMS. El objetivo de este instrumento es captar la CVRS, considerando la percepción de las personas en su contexto cultural, su entorno, su sistema de valores, metas personales, estándares y preocupaciones. Aparte de la versión extensa inicial de 100 ítems, cuenta con una versión abreviada de 26 ítems, una adaptación para pacientes con SIDA y una serie de preguntas adicionales que evalúan aspectos de espiritualidad, religiosidad y creencias personales (World Health Organization 2017). Finalmente, el EuroQol (EQ-5D) aproxima la CVRS basado en los ítems de movilidad, cuidado personal, actividades diarias, dolor o malestar, ansiedad o depresión, y salud general. Esta escala se implementa habitualmente en estudios de costes e uso de recursos (Herdman et al. 2001), así como en estudios que cuantifican el impacto de la enfermedad en términos de años de vida ajustados por la calidad (QALY) (Seidl et al. 2017, Mangen et al. 2017).

A pesar de su gran utilidad y extendido uso, desde el punto de vista psicométrico, los cuestionarios genéricos se consideran poco sensibles a los cambios de salud debidos a patologías concretas de los individuos. Esta función se debería cumplir por las escalas denominadas específicas (Coons et al. 2000, Ware, Jr. et al. 1998).

### **Escalas específicas**

Las escalas específicas están diseñadas para emplearse en la evaluación del estado de salud de los sujetos que padecen patologías concretas, como pueden ser por ejemplo los pacientes pediátricos con asma (Liu et al. 2007), mujeres en menopausia (Hilditch et al. 1996), pacientes con cáncer (Aronson et al. 1993), individuos con obesidad mórbida (Karlsson et al. 2003), o pacientes con patologías de la articulación del hombro (Michener & Leggin 2001). Este tipo de cuestionarios suele evaluar, entre otros aspectos, la presencia y el nivel del dolor; las limitaciones que puede causar la enfermedad en las actividades de ocio, del trabajo y de la vida diaria en general; así como el funcionamiento psicológico, la imagen corporal o las conductas sociales de las personas. Debido a este enfoque, las escalas específicas son más sensibles a los cambios producidos a lo largo del tiempo, mientras que también tienen una mayor capacidad de discriminación entre diferentes grados de la enfermedad (Mc Dowell 2006).

### **Escalas específicas para la articulación del hombro**

En el ámbito de la patología del hombro, existen más de 20 escalas auto-cumplimentadas para la evaluación de esta articulación. En la Tabla 3.1 se presentan las más relevantes, que se podrían dividir en tres grupos, las escalas creadas para evaluar el miembro superior en general (mano, codo, hombro), las que evalúan el hombro con independencia de su patología y las diseñadas para patologías concretas de la articulación (Bot et al. 2004, Roe et al. 2013, Giesinger et al. 2015).

Con relación al primer grupo, el instrumento más referenciado es el DASH (Hudak et al. 1996) y su versión abreviada, el QuickDASH (Beaton et al. 2005). Entre las escalas creadas para la valoración del hombro, independientemente de su patología, las más utilizadas son las ASES-p, SPADI, SST y OSS (Roe et al. 2013, Schmidt et al. 2014). Las dos primeras, aparte de una puntuación total, cuentan además con dos sub-escalas las de *dolor* y *función*. Cabe mencionar que el cuestionario UCLA también se encuentra entre los más referenciados de este grupo. Aunque 2 de sus 5 ítems no son auto-cumplimentados (flexión y fuerza), los otros 3 (dolor, función y satisfacción) se suele

utilizan como escala independientemente, mientras que en el año 2008 un grupo de investigadores australianos presentó una versión de UCLA completamente auto-administrada (Coghlan et al. 2008). Entre los cuestionarios creados para patologías concretas de hombro, los más referenciados son los WOSI y WORC diseñados para las patologías de la inestabilidad y del manguito rotador (Roe et al. 2013). Ambas escalas fueron diseñadas por la universidad de Western Ontario (Canadá) y exploran varias sub-dimensiones como son los síntomas físicos, la actividad, el tiempo libre, y las emociones. Por los cuestionarios de este último grupo, se aprecia el interés y la relevancia que tienen para los investigadores las dos patologías arriba mencionadas. Tres de las seis escalas de este grupo evalúan la inestabilidad del hombro y otras dos el manguito rotador.

Es evidente, que entre las escalas de este ámbito existen muchas diferencias. Entre ellas, el número total de ítems, las dimensiones de interés, o el tipo de preguntas destinadas a cada una de ellas, en definitiva, no existen dos cuestionarios iguales. Aun así, el objetivo final de todas ellas es realizar evaluaciones sobre el estado de la articulación, en base a una serie de aspectos como las AVD, el dolor, la funcionalidad, la actividad física, el trabajo, e incluso las emociones de los individuos; todo ello desde la perspectiva *subjetiva* de los pacientes. En la tesis actual no se realiza una presentación exhaustiva de todas las escalas auto-cumplimentadas que se utilizan en la patología del hombro. Aun así, las aquí presentadas estarían entre las más relevantes en este ámbito (Longo et al. 2011, Roe et al. 2013, Plancher & Lipnick 2009, Rouleau et al. 2010, Baumgarten & Helsper 2016).

**Tabla 3.1:** Escalas auto-cumplimentadas utilizadas en la evaluación de hombro

Escala	Nombre extenso	Autor-año	Nº ítems	Dimensiones
<b>Extremidad superior</b>				
DASH	Disabilities of the Arm, Shoulder and Hand Scale	Hudak-1996	30	-
QuickDASH	Version abreviada de DASH	Beaton-2005	11	-
UEFS	Upper Extremity Function Scale	Pransky-1997	8	-
UEFLS	Upper Extremity Functional Limitation Scale	Simonsick-2001	7	-
UEFI	Upper Extremity Functional Index	Stratford-2001	20	-
ULFI	Upper Limb Functional Index	Gabel-2006	25	-
<b>Articulación del hombro en general</b>				
UCLA	University of California Los Angeles Shoulder Score, self-evaluation section	Ellman-1986	3	Dolor, función, satisfacción
SSI	Shoulder Severity Index	Patte-1987	30	-
SPADI	Shoulder Pain and Disability Index	Roach-1991	13	Dolor, función
SST	Simple Shoulder Test	Lippitt-1993	12	-
SDQ-UK	United Kingdom Shoulder Disability Questionnaire	Croft-1994	22	-
ASES-p	American Shoulder and Elbow Surgeons Shoulder Assessment Form, self-evaluation section	Richards-1994	11	Dolor, función
OSS	Oxford Shoulder Score	Dawson-1996	12	-
SPS	Shoulder Pain Score	Winters-1996	6	-
SRQ	Shoulder Rating Questionnaire	L'Insalata-1997	21	Valoración general, dolor, AVD, trabajo, tiempo libre, satisfacción, mejora
SSRS	Subjective Shoulder Rating Scale	Kohn-1997	5	-
PSS	Penn Shoulder Score	Leggin-1999	24	Dolor, función, satisfacción
SDQ-NL	Dutch Shoulder Disability Questionnaire	van der Heijden-2000	16	-
FLEX-SF	Flexilevel Scale of Shoulder Function	Cook-2003	15	-
<b>Patologías concretas del hombro</b>				
WOSI	Western Ontario Shoulder Instability Index	Kirkley-1998	21	Síntomas, AF, tiempo libre y trabajo, estilo de vida, emociones
OSI	Oxford Shoulder Instability	Dawson-1999	12	-
MISS	Melbourne Instability Shoulder Score	Watson-2005	22	Dolor, inestabilidad, función, AF y trabajo
RC-QOL	Rotator Cuff Quality of Life Measure	Hollinshead-2000	34	-
WOOS	Western Ontario Osteoarthritis of the Shoulder Index	Lo-2001	19	Síntomas, AF, tiempo libre y trabajo, estilo de vida, emociones
WORC	Western Ontario Rotator Cuff Index	Kirkley-2003	21	Síntomas, AF y tiempo libre, trabajo, estilo de vida, emociones

Dimensiones: Se indican las diferentes dimensiones y el número de sus correspondientes ítems (en paréntesis); el signo (-) indica la existencia de una única dimensión. Todas las escalas cuentan con una puntuación total, independientemente de las sub-dimensiones de las mismas. AVD: actividades de la vida diaria. AF: actividad física.

### **3.2. Capacidad funcional basada en la movilidad y la observación clínica**

El término *función* engloba diferentes conceptos, como la capacidad física, capacidades cognitivas, sensoriales, psíquicas y sociales. Como se expone en la sección anterior, la *función*, es una parte integral de la CVRS. A su vez, la movilidad es una parte importante de la funcionalidad. Existen varias pruebas diseñadas para evaluar aspectos de la movilidad de los individuos. En algunos casos, estos aspectos se valoran mediante preguntas auto-cumplimentadas. En otros, la evaluación se basa exclusivamente en la observación clínica, mientras que también existen pruebas y escalas que combinan ambos formatos de evaluación.

Aunque los conceptos de la *función* y *prueba funcional* son muy amplios y se pueden utilizar para referirse a diferentes tipos de pruebas y escalas, las pruebas de movilidad basadas en la observación clínica se denominan a menudo *pruebas funcionales* o *pruebas de rendimiento*. Este tipo de pruebas se compone de una serie de acciones estandarizadas, cuyos resultados se recogen por un evaluador externo. El peso que la observación clínica tiene en estos casos y la ausencia de la auto-percepción del paciente, hace que los respectivos resultados se perciban como *objetivos*. Estas pruebas son ampliamente utilizadas, sobre todo en estudios de fragilidad, de riesgo de caídas y de discapacidad (Newman & Cauley 2012, Gill 2010, Lusardi et al. 2017).

#### **3.2.1 Pruebas funcionales de observación clínica**

Las *pruebas funcionales*, basadas en la observación clínica, se pueden separar en dos grupos principales, las pruebas que evalúan la movilidad de las extremidades inferiores y las pruebas basadas en la movilidad de las extremidades superiores (Newman & Cauley 2012).



### ***Extremidad inferior***

En relación a la extremidad inferior, las pruebas funcionales más utilizadas son el *levántate y anda cronometrado* (LAC) y la *velocidad de la marcha* (VM). El LAC mide el tiempo que tardan los sujetos en levantarse de una silla, caminar 3 metros, girar, regresar a la silla y sentarse de nuevo (Podsiadlo & Richardson 1991). La VM mide los metros por segundo que hacen los pacientes, cuando recorren a su ritmo habitual, una distancia de 4 metros (Guralnik et al. 2000). Por otro lado, el *Short Physical Performance Battery* (SPPB) es una escala más extensa, que incluye la VM, una serie de pruebas de equilibrio y la capacidad de levantarse y sentarse varias veces de una silla. Aunque estas pruebas se usan en el ámbito de la fragilidad y la disminución de la funcionalidad, en personas mayores, también se aplican a pacientes con otras patologías, como artrosis de la rodilla o enfermedad degenerativa del disco lumbar (Gill 2010, Corniola et al. 2016, Dere et al. 2014). Además el SPPB se ha utilizado en pacientes en hemodiálisis y en sujetos con enfermedad renal (Lorenz et al. 2017, Abreo et al. 2014).

### ***Extremidad superior y la articulación del hombro***

Escribir una frase, coger objetos pequeños, abotonar una camisa, colocar pernos en un tablero (Pegboard test) o la fuerza de la mano medida con un dinamómetro, son algunas de las pruebas que se utilizan para estudiar la funcionalidad de la extremidad superior en general (Gill 2010, Bohannon 2008, Hazuda et al. 2005). En este caso, destacar las más utilizadas entre ellas no es una tarea fácil, ya que en la literatura científica no se encuentra el mismo soporte de evidencia, que con las pruebas para la extremidad inferior (Gill 2010, Newman & Cauley 2012).

Aun así, cuando nos centramos en la articulación del hombro, como *funcionalidad objetiva* se consideran el rango de movimiento y la fuerza de la articulación. Estas mediciones se realizan con la ayuda de goniómetros y tensiómetros y son parámetros muy relevantes para los clínicos (Fong et al. 2013, Baldwin et al. 2013, Vincent et al. 2016, Cvetanovich et al. 2015, Baumgarten & Helsper 2016). En 1987 los traumatólogos Constant y Murley presentaron una escala, la denominada Constant-Murley Score

(CMS), que combinaba la evaluación del rango del movimiento y de la fuerza con preguntas subjetivas sobre dolor y AVD. A pesar de los componentes *subjetivos* de este cuestionario la mayor parte del mismo ( 65% de su puntuación total) requiere la participación y evaluación de un especialista traumatólogo. Principalmente por este motivo, a lo largo de los años el CMS se estableció como regla de oro, para muchos profesionales e investigadores, en el estudio de esta articulación (Barra-Lopez 2007, Roy et al. 2010, Constant et al. 2008). Debido al extendido uso, interés y relevancia de esta escala en la investigación de la articulación del hombro, uno de los estudios de la presente tesis está dedicada a la valoración y evaluación de las propiedades de la misma. Información detallada sobre la escala y datos adicionales relacionados con su funcionamiento en diferentes patologías del hombro, se pueden encontrar en los capítulos 4 y 8 respectivamente.

### 3.3 Atributos y propiedades psicométricas

La psicometría es la ciencia relacionada con la construcción de instrumentos de medición y la evaluación de si estos instrumentos son válidos y fiables para medir lo que pretenden medir. Por lo general, en el ámbito de la salud, la medición se realiza mediante cuestionarios que deben evaluarse ampliamente, antes de poder afirmar que poseen buenas propiedades psicométricas (Gellman & Turner 2013).

El *Medical Outcomes Trust* (MOT) es una organización estadounidense, sin ánimo de lucro, cuyo principal objetivo es proporcionar guías de control de calidad para este tipo de escalas de medición. En el año 1996, el *Scientific Advisory Committee* (SAC) del MOT publicó un artículo sobre criterios de revisión de los instrumentos que incorporan las perspectivas de los pacientes (Lohr et al. 1996). La evolución experimentada en este ámbito y las nuevas evidencias, pronto condujeron a la re-evaluación y actualización de estos criterios. Como resultado, en 2002 el SAC publicó en la revista *Quality of Life Research* su segundo y más referenciado artículo sobre las características métricas que se tienen que considerar en las evaluaciones de los instrumentos de CVRS (Aaronson et al. 2002). Estas fueron las siguientes:

1. Modelo conceptual y de medida
2. Interpretabilidad
3. Carga de administración y de respuesta
4. Formas alternativas de administración
5. Adaptación cultural y del lenguaje
6. Fiabilidad
7. Validez
8. Sensibilidad al cambio

A lo largo de los años, estas características se han ido consolidando (Terwee et al. 2007, Valderas et al. 2008, Valderas et al. 2005, Mokkink et al. 2010, Reeve et al. 2013). Aun así, cabe subrayar que las propiedades de fiabilidad, validez y sensibilidad al cambio, son las más importantes en el ámbito de las escalas de CVRS. El hecho de que existan

métodos estadísticos concretos para la medición de las mismas, implica que su estudio y evaluación siguen ciertos pasos estandarizados. Eso hace que estas tres cualidades obtengan mayor atención que el resto. La decisión de los investigadores sobre si utilizar o no una escala se basa a menudo en estas tres propiedades, de manera individual o conjuntamente, según el objetivo de la investigación (van de Water et al. 2014, Staples et al. 2010, Mc Dowell 2006, Sijtsma 2009, Morkink et al. 2010, Mc Dowell 2006, Zumbo & Chan 2014).

En las siguientes secciones se presenta información más detallada sobre las 8 características del MOT. En este punto, es importante mencionar que organismos internacionales como el *U.S Food and Drug Administration*, *American Psychological Association*, y *European Federation of Psychologists' Association*, entre otras, trabajan y publican periódicamente directrices y criterios requeridos para las escalas de CVRS, similares a los aquí presentados (Zumbo & Chan 2014).

### **3.3.1 Modelo conceptual y de medida**

Se refiere a la justificación y descripción del concepto que un instrumento pretende evaluar, por ejemplo: ansiedad, depresión, afrontamiento etc. El concepto tiene que estar bien argumentado, mientras que los diferentes ítems y componentes de la escala tienen que contar con evidencia conceptual y empírica. Es importante la participación de la población diana y de expertos en el campo objeto del estudio durante la creación del instrumento. Se tiene que proporcionar información precisa sobre el nivel de medida, la puntuación total y su variabilidad, así como sobre las sub-dimensiones de la escala, si existen, justificación sobre las mismas y los ítems que corresponden a cada una de ellas (Aaronson et al. 2002, Valderas et al. 2005).

### **3.3.2 Interpretabilidad**

Es la capacidad de asignar un significado comprensible a las puntuaciones de la escala. Esta capacidad se facilita si existen *puntuaciones de referencia*, basadas en poblaciones

sanas o puntuaciones entre grupos de distintas características como por ejemplo: presencia/ausencia de una patología, diferentes grupos de gravedad, tratamientos de distinta efectividad. Establecer los valores mínimos clínicamente importantes para los pacientes (*minimal important change*) permite una mejor interpretación de los cambios observados a lo largo del tiempo. Finalmente, la capacidad de la herramienta de predecir eventos externos como pueden ser la institucionalización o la muerte, facilita y ofrece una perspectiva más clara sobre la interpretabilidad de sus puntuaciones (Aaronson et al. 2002, Mokkink et al. 2010, Valderas et al. 2005)

### 3.3.3 Facilidad de uso

Este atributo hace referencia al tiempo requerido y los recursos necesarios para la administración de una escala, tanto por parte del entrevistador como por parte del entrevistado.

Más en concreto, por parte del entrevistador, se evalúan los medios necesarios para la administración de un cuestionario. Entre otros, se considera si se realiza cara a cara, por teléfono o por ordenador, si requiere software específico o si precisa de manual. También se evalúan la necesidad de consultar documentos adicionales, como por ejemplo la historia clínica o información específica sobre los medicamentos que utilizan los participantes. Asimismo, se evalúa el nivel de conocimiento específico que necesita el entrevistador para poder realizar una administración efectiva de la escala.

Por parte del entrevistado, sobre todo en los casos de cuestionario auto-administrados, se evalúa la capacidad de lectura y comprensión requeridas, la adaptación del diseño de impresión a las necesidades del grupo poblacional de interés (tipo y tamaño de letra, formato de opción de respuestas, espacios entre preguntas etc.). Otro aspecto relevante tiene que ver con la aceptabilidad de las preguntas y la posible tensión física o emocional, que pueden suponer para el encuestado (Aaronson et al. 2002, Valderas et al. 2005).

### 3.3.4 Formas alternativas de administración

Se considera como forma alternativa de administración cualquier formato adicional o diferente al originalmente diseñado para la administración de un instrumento de CVRS. Por ejemplo, para instrumentos administrados por un especialista, formas alternativas serían adaptaciones auto-administradas, o incluso instrumentos adaptados para familiares o proxies de los sujetos reclutados. Como parte de este atributo, se tiene que evaluar toda la información disponible relacionada con las propiedades de fiabilidad, validez, sensibilidad al cambio, interpretabilidad y facilidad de uso de las formas alternativas de administración. Asimismo, se tiene que proporcionar información sobre la comparabilidad entre las diferentes formas de administración de una escala (Aaronson et al. 2002, Valderas et al. 2005).

### 3.3.5 Proceso de adaptación cultural & lingüística

Este atributo se evalúa en los casos de las escalas adaptadas a otro idioma, distinto a su idioma original. Se valora la equivalencia conceptual y lingüística de la escala en el nuevo idioma, así como sus propiedades métricas. En relación a la equivalencia lingüística el SAC aconseja seguir los siguientes pasos: 1) realizar dos traducciones independientes desde el idioma original al idioma diana; 2) a continuación realizar una o más traducciones desde el idioma diana al idioma original; 3) revisión de las versiones retro-traducidas por panel de expertos; 4) proporcionar evidencia sobre aceptabilidad, comprensión y equivalencia de la versión adaptada (Aaronson et al. 2002, Valderas et al. 2005).

### 3.3.6 Fiabilidad

La fiabilidad se define como la ausencia de error aleatorio de una escala. Como parte de este atributo se evalúa la *consistencia interna* y la *reproducibilidad*. La *consistencia interna* hace referencia a la homogeneidad de los ítems de un cuestionario. Esta condición es primordial para cualquier escala de medición, ya que las preguntas que

componen una misma escala o sub-escala en el caso de que exista, tienen que medir el mismo concepto. Eso significa que tienen que presentar cierta relación entre ellos. Para los cuestionarios de ítems ordinales, tipo Likert, esta propiedad se evalúa sobre todo, mediante el coeficiente alpha de Cronbach, mientras que en los casos de ítems binarios se cuantifica con el coeficiente de Kuder-Richardson (Bland & Altman 1997). Los modelos basados en la teoría de respuesta al ítem estiman la fiabilidad marginal de los ítems, como medida correspondiente al alpha de Cronbach.

La *reproducibilidad* se refiere a la estabilidad de respuestas del instrumento. Dicha estabilidad se puede medir tanto en los mismos sujetos en dos momentos en el tiempo (test-retest) o entre entrevistadores en un mismo momento. En el caso de la reproducibilidad test-retest es importante tanto la justificación del intervalo entre mediciones, como la estabilidad de los sujetos en el tiempo transcurrido. Esta estabilidad se refiere a la ausencia de cambios clínicos o de otros aspectos, que puedan haber modificado las puntuaciones. Por otro lado, cuando se comprueba la concordancia entre diferentes entrevistadores, es importante que ellos estén en igualdad de condiciones en cuanto a la administración de la escala se refiere (i.e. mismo nivel de experiencia, formación sobre el uso del instrumento). Existen diferentes métodos para comprobar la reproducibilidad de las escalas. El coeficiente de correlación Intraclass (CCI) y los límites de confianza al 95% de Bland & Altman se utilizan para comprobar la reproducibilidad de las escalas continuas. El coeficiente de Kappa se implementa en las escalas ordinales. Los modelos de la teoría de respuesta al ítem proporcionan información sobre la comparabilidad de las estimaciones de los ítems y su precisión en repetidas administraciones (Aaronson et al. 2002, Valderas et al. 2005, Terwee et al. 2007, Mc Dowell 2006).

### **3.3.7 Validez**

La validez se define como el grado en el que un instrumento consigue medir lo que pretende medir. Es quizás una de las propiedades más importante en el ámbito de la validación de las escalas y la que más evolución, al menos en términos de definiciones conceptuales, ha experimentado a lo largo de los años (Anastasi 1986, Guion 1980,

CRONBACH & MEEHL 1955, Messick 1995). Los tres principales tipos de validez son, la validez de *contenido*, la validez de *constructo* y la validez de *criterio*.

Con el término *validez de contenido* nos referimos a la evidencia sobre el contenido apropiado y relacionado con el concepto que el instrumento pretende medir. Dicha validez se obtiene con la participación de expertos y sujetos de la población de interés, en el proceso de la elaboración y selección de las preguntas de un cuestionario, así como mediante revisiones de la literatura sobre los conceptos relevantes. La claridad de las preguntas, la comprensión de las mismas, por parte de los participantes, y la posible redundancia de los ítems o de las sub-escalas de un cuestionario, son también parte de este tipo de validez. La validez de contenido no se suele evaluar de manera cuantitativa, ya que es muy difícil, si no imposible, demostrar que todos los elementos pertinentes están debidamente representados por los ítems de una escala (Mc Dowell 2006, Terwee et al. 2007, Valderas et al. 2005).

La *validez de constructo* hace referencia al grado de correspondencia entre los componentes de un instrumento y el constructo que este instrumento pretende medir, dando soporte a la interpretación de las puntuaciones del instrumento. En este tipo de validez cuenta con varios aspectos y consecuentemente con una amplitud de métodos para la evaluación de los mismos. La *validez estructural*, la *validez convergente* y *discriminante*, y la *validez de grupos conocidos* son parte de la validez de constructo (Mokkink et al. 2010, Mc Dowell 2006).

La *validez estructural* tiene que ver con la evaluación de la dimensionalidad de una escala y si esta refleja adecuadamente el constructo que pretende medir. Este constructo hace referencia al factor o factores latentes (no observables) que cada instrumento explora. Los modelos de análisis factorial confirmatorio (Brown 2006) y el modelo Rasch (Bond & Fox 2007) son algunas de las herramientas que permiten explorar la validez de estructura y comprobar la adecuación del conjunto de los ítems de la escala en relación a los factores latentes de interés. La *validez convergente* y *discriminante* estudia la correlación de un cuestionario con otros que miden constructos similares y diferentes respectivamente. Se espera que las correlaciones entre



instrumentos de constructos parecidos sean elevadas, mientras que lo contrario debería suceder entre escalas de constructos diferentes. Finalmente, la *validez de grupos conocidos* se comprueba mediante el estudio de las puntuaciones totales de la escala en distintos grupos de pacientes, que deberían presentar puntuaciones diferentes. Estos grupos se pueden definir por niveles de gravedad, grupos de edades, niveles funcionales etc. (Terwee et al. 2007, Mc Dowell 2006).

Por último, con *validez de criterio* nos referimos a la relación del instrumento con otra medida, que se considere un estándar de oro en el respectivo ámbito de investigación (*gold standard*). Este tipo de validez no se suele evaluar en el caso de las medidas de CVRS, ya que no suele existir una medida ampliamente aceptable que se utilice como referencia para este tipo de escalas. Aun así, cabe mencionar que cuando se evalúan versiones cortas de una escala de CVRS, como estándar de oro se pueden considerar las versiones más extensas de la misma (Aronson et al. 2002, Valderas et al. 2005).

### **3.3.8 Sensibilidad al cambio**

Es la capacidad del instrumento para detectar cambios sucedidos en el tiempo. Se evalúan las puntuaciones de la escala de CVRS observadas a lo largo del tiempo para pacientes estables o pacientes cuyo estado ha experimentado cambios. La magnitud de los efectos observados entre medidas consecutivas tiene que estar acorde con las características y la evolución del estado de los pacientes. Dichos cambios se miden con estadísticos que transforman las diferencias pre-post a unidades estandarizadas en términos de la varianza, como son el tamaño del efecto o el error estándar de la medida (Aronson et al. 2002, Valderas et al. 2005, Husted et al. 2000, Mc Dowell 2006).

### **3.4 Herramientas de evaluación de las escalas**

En el ámbito de la salud existen varias iniciativas cuyo objetivo es apoyar la selección de las escalas de calidad de vida más adecuadas (Bombardier & Tugwell 1987, Andresen 2000, Streiner 1993, Badia & Baro 2001, Valderas et al. 2008, Mokkink et al. 2012, DeVon et al. 2007). En la sección actual se expone información sobre las más recientes entre ellas: el índice GraQol (Badia & Baro 2001) y las herramientas EMPRO (Valderas et al. 2008) y COSMIN (Mokkink et al. 2012). Las dos primeras proporcionan como resultado final una puntuación global que permite ordenar los cuestionarios según su desarrollo y calidad, respectivamente, mientras que la tercera no realiza una síntesis a partir de la evaluación de las diferentes propiedades.

#### **3.4.1 GraQol**

La creación del índice GraQol fue una iniciativa de investigadores españoles (Badia & Baro 2001). Este índice, se basa en la evaluación de 11 criterios relacionados con la adaptación al idioma diana y la prueba piloto del cuestionario adaptado, la validación estructural, convergente, y de los puntos de corte. También evalúa la fiabilidad y la sensibilidad al cambio (Tabla 3.2). Para cada criterio evaluado, se reciben 2 puntos, mientras que a los criterios no evaluados se asignan 0 puntos. La puntuación total del índice es el valor porcentual de la suma de los puntos en los criterios precedentes, entre la máxima puntuación posible, considerando solo los criterios relevantes. Es decir, para un cuestionario creado y evaluado en España no se consideraría en el cálculo de este índice los criterios de “1.traducción y retro-traducción” y de la “2.prueba piloto de la adaptación”, ambos relacionados con la adaptación cultural.

Según sus autores, el GraQol se creó como una herramienta descriptiva sobre el grado de desarrollo de un cuestionario de CVRS y especialmente para los cuestionarios utilizados en el ámbito de la atención primaria. Según las indicaciones proporcionadas, una puntuación total <50% mostraría escalas con un desarrollo pobre, valores entre 50-70% implicarían un desarrollo aceptable, mientras que valores >70% sugerirían un desarrollo satisfactorio del cuestionario de interés. Aunque este índice fue una iniciativa

interesante, su uso no se ha extendido. La falta de instrucciones concretas sobre la correcta cumplimentación de sus componentes, así como su enfoque en el ámbito de la atención primaria o incluso la definición de herramienta “descriptiva”, podrían ser unos posibles motivos de esta falta de trascendencia (Valderas 2008).

**Tabla 3.2:** Índice GraQol para la evaluación de cuestionarios

---

1. Traducción y retro traducción
2. Prueba piloto de la adaptación
3. Validación adaptación de las ponderaciones
4. Validación estructural
5. Validación convergente
6. Validación/adaptación de los puntos de corte
7. Sensibilidad en distintas poblaciones-validez discriminantes
8. Fiabilidad interna
9. Fiabilidad test-retest
10. Fiabilidad entre observadores
11. Sensibilidad al cambio

---

Sistema de evaluación de los criterios del índice:  
NP: no procede; 0: no se ha efectuado/no se conoce y 2: se ha efectuado

$$\text{Índice GraQol} = \frac{\text{Suma de la puntuación real en los criterios procedentes}}{\text{Máxima puntuación posible entre los criterios procedentes}} \times 100$$

---

Referencia: X. Badia, E. Baró; Atención primaria 2001; 28:349-356

### 3.4.2 EMPRO

La herramienta *Evaluating the Measurement of Patient-Reported Outcomes* (EMPRO), es un instrumento estandarizado, cuyo objetivo es valorar las propiedades psicométricas de los cuestionarios de CVRS (Valderas et al. 2008). EMPRO se desarrolló en el entorno de la Red de Investigación en Resultados Y Servicios de Salud (IRYSS) a partir de las recomendaciones realizadas por el Scientific Advisory Committee (SAC) del Medical Outcomes Trust (MOT). Consiste de 39 ítems divididos en los 8 atributos que propone el SAC del MOT (Aronson et al. 2002). Estos atributos son: *concepto y modelo de medida, fiabilidad, validez, sensibilidad al cambio, interpretabilidad, facilidad de uso del instrumento, formas alternativas de administración, proceso de adaptación cultural y*

lingüística. Cada atributo se compone de 2 a 8 ítems. Esta herramienta evalúa conjuntamente toda la información disponible en la literatura sobre un mismo cuestionario de CVRS. Las puntuaciones de los ítems individuales se combinan, para proporcionar una puntuación por atributo y una puntuación global, en base a los primeros 5 atributos arriba mencionados (Figura 3.1). Todas las puntuaciones oscilan de 0-100 puntos, representando el mínimo y máximo valor respectivamente. Puntuaciones aceptables se consideran aquellas que llegan por los menos a los 50 puntos. El EMPRO es una herramienta válida y fiable que se ha implementado para la evaluación de cuestionarios de CVRS tanto específicos como genéricos (Garin et al. 2013, Schmidt et al. 2014, Valderas et al. 2008, Maratia et al. 2016). El EMPRO se ha utilizado en dos estudios de la presente tesis. Información más extensa sobre él se presenta en el capítulo 4, mientras que la escala se encuentra en el Anexo I.

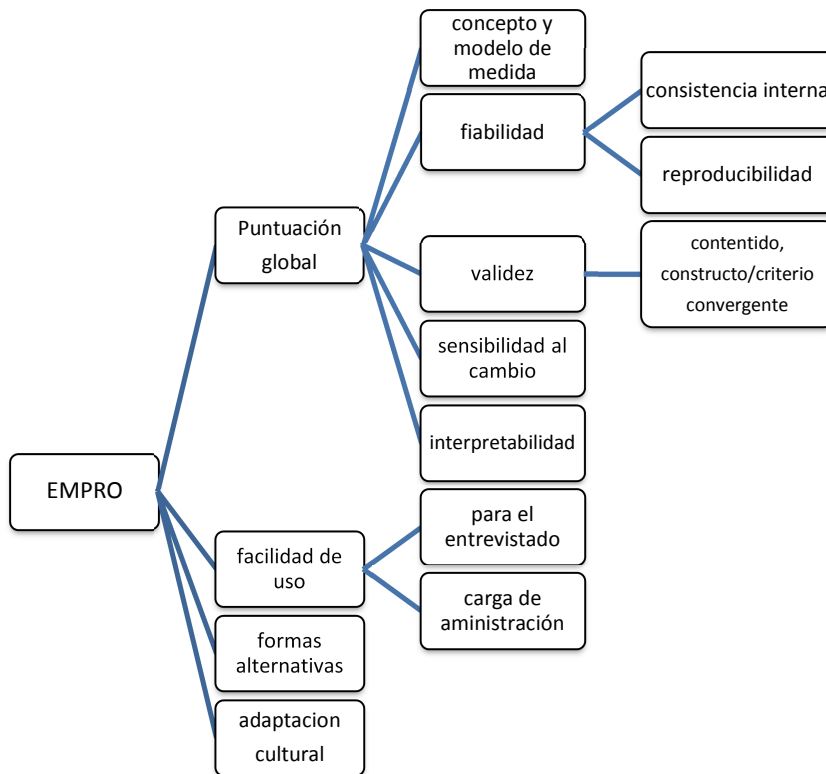


Figura 3.1: Propiedades métricas evaluadas por la herramienta EMPRO.

### 3.4.3 COSMIN

La herramienta *Consensus based Standards for the Selection of health status Measurement Instruments* (COSMIN), fue creada a través de un estudio Delphi internacional, liderado por un grupo Holandés del University Medical Centre de Ámsterdam y es posterior a EMPRO (Mokkink et al. 2010). Se compone de 119 ítems divididos en los siguientes 8 atributos: *consistencia Interna, fiabilidad, error de medición, validez de contenido, validez estructural, evaluación de hipótesis, validez de la adaptación cultural, validez de criterio, sensibilidad al cambio, e interpretabilidad*. La taxonomía de las propiedades de COSMIN se presenta en la Figura 3.2.



**Figura 3.2:** La taxonomía de las propiedades métricas. Imagen copiada del manual de COSMIN (<http://www.cosmin.nl>).

Originalmente los ítems de COSMIN se puntuaban con dos alternativas de respuesta: Si/No (Mokkink et al. 2010). En una publicación posterior, los autores presentaron esta misma escala con 4 alternativas de respuesta, para obtener mayor variabilidad en las puntuaciones finales (Terwee et al. 2012).

En su mayoría, los componentes de COSMIN se asemejan a los de EMPRO, aunque no existe una correspondencia exacta entre todos ellos. También coincide el objetivo final de las dos herramientas, que es facilitar la selección del cuestionario más apropiado en cada contexto. Aun así, existe una diferencia sustancial entre ellas. COSMIN fue creado para evaluar la calidad metodológica de estudios individuales sobre las cualidades psicométricas de un cuestionario. Eso, por lo menos en teoría, significa que cada artículo científico se tiene que evaluar por separado y que las distintas evaluaciones no se combinan, tarea que sí se realiza con EMPRO. Desde su creación, el COSMIN también se ha implementado en varios estudios científicos (Lee et al. 2016, Kotecha et al. 2016, Uijen et al. 2012, Cheong & Johnston 2013). Aunque en muchos casos, no queda claro cómo exactamente se realiza la síntesis de la evidencia. Esta herramienta no se utilizó en ninguno de los estudios de la presente tesis, principalmente debido a la falta de una puntuación global y las dificultades que eso implica a la hora de realizar comparaciones entre varias escalas.

## **4. Material y Métodos**

En este capítulo se presentan los materiales y métodos implementados en los diferentes estudios que componen la presente tesis doctoral. Dado que cada uno de estos estudios se expone más adelante en formato de artículo científico (capítulos 5-8), cierta información aquí incluida puede verse repetida con más detalle en los correspondientes capítulos.

## **4.1 Estudio de evaluación de las escalas auto-cumplimentadas**

Información detallada sobre este estudio, así como los resultados del mismo se presentan en el capítulo 5.

### ***Revisión sistemática y evaluación de la evidencia***

El objetivo de este trabajo, fue realizar una evaluación estandarizada de las escalas de CVRS auto-cumplimentadas, para pacientes con patologías del hombro, con objetivo final establecer el cuestionario más adecuado para su posterior adaptación y validación al idioma diana. Se realizó una búsqueda sistemática en la biblioteca virtual de PubMed para identificar cuestionarios multi-ítem e auto-cumplimentados, desarrollados para pacientes con patologías del hombro. La búsqueda se limitó a los artículos publicados en inglés. Las cualidades psicométricas de todos los cuestionarios identificados se evaluaron de manera independientes por dos revisores expertos, mediante la herramienta EMPRO (sección 4.4).

## **4.2 Estudio de validación al castellano de la escala ASES-p**

Información detallada sobre el estudio de la validación de la escala ASES-p y los resultados del mismo se presentan en el capítulo 6. Datos adicionales, sobre la fiabilidad test-retest del cuestionario se presentan en el capítulo 7.

Los resultados del estudio anterior (sección 4.1), indicaron que de manera global, la escala ASES-p contaba con mejores cualidades psicométricas que el resto de las escalas evaluadas. El ASES-p se adaptó culturalmente y, tras un estudio piloto, se administró a una muestra de sujetos con patologías de hombro, para obtener los datos necesarios para su validación al castellano.



### ***Adaptación cultural de ASES-p***

La adaptación cultural de la escala ASES-p se realizó siguiendo los pasos indicados en el protocolo IQOLA (Bullinger et al. 1998). Un traumatólogo y un traductor profesional, con idioma materno el castellano, realizaron de manera independiente una traducción de la escala, del inglés al castellano. El traductor profesional desconocía el tema del proyecto y no estaba familiarizado con las patologías de la articulación del hombro. A continuación, ambos traductores recibieron la otra traducción. Las dos partes debatieron sobre las diferentes opciones de traducción al idioma diana y consensuaron la primera versión castellana de ASES-p. En la siguiente fase, dos traductores cuyo idioma materno era el inglés, con fluidez al castellano, tradujeron la primera versión castellana de ASES-p al inglés (back-translation). Esta última versión traducida se comparó con la versión inglesa de ASES-p. Se debatieron y se resolvieron las discrepancias entre ambas partes, realizando los cambios pertinentes en la versión castellana. Un comité, compuesto por dos traumatólogos expertos en la extremidad superior, un epidemiólogo con larga experiencia en medidas de CVRS, un traductor profesional y la autora del presente trabajo aceptaron la versión traducida de ASES-p.

Entre los 11 ítems de la escala el único que presento cierta dificultad a la hora de adaptarlo al castellano fue el ítem 4 “Manage toileting”. Uno de los traductores consideró que esta pregunta estaba relacionada con actos que tienen que ver con el cuidado de la apariencia, mientras que el otro lo percibió como la dificultad que pueden tener los sujetos a la hora de limpiarse tras la micción y defecación. Para aclarar cuál de las dos opciones era la más apropiada, nos pusimos en contacto con la autora principal de la validación estadounidense de ASES-p (Michener et al. 2002). Ella aclaró que se trataba de la segunda opción arriba mencionada. A continuación se consensó la primera versión castellana de ASES-p, esta se tradujo de nuevo al inglés y se comprobó que no hubo diferencias conceptuales entre la versión original y la retro traducida.

### ***Estudio piloto***

A continuación la primera versión castellana se administró a una muestra de 10 sujetos con patologías de hombro, elegidos de manera consecutiva en las consultas de traumatología de los dos hospitales participantes. Dos de los sujetos no contestaron al ítem sobre la dificultad en realizar su deporte habitual, ya que no hacían ningún tipo de deporte. El resto de las preguntas se rellenaron por todos los sujetos que consideraron que la versión adaptada de ASES-p era comprensible y fácil de rellenar. Además, todos los ítems se consideraron relevantes con la patología de interés. Por lo tanto no se realizaron más cambios a esta versión de la escala. Los participantes del piloto no fueron reclutados para la validación cultural de la escala.

### ***Sujetos y centros participantes***

Para realizar la validación cultural de la escala se reclutaron de manera prospectiva sujetos  $\geq 18$  años, en lista de espera para recibir una intervención quirúrgica o conservadora en la articulación de hombro. El reclutamiento se realizó en los servicios de traumatología y cirugía ortopédica de cinco centros hospitalarios de Osakidetza, en los Hospitales Universitarios Donostia, Basurto, Cruces, Hospital Galdakao-Usansolo y Hospital Mendaró. El reclutamiento se realizó entre mayo 2012 y noviembre 2013.

### ***Recogida de datos***

Los traumatólogos colaboradores cumplimentaron la herramienta funcional de Constant-Murley Score (CMS) (Constant & Murley 1987). Los sujetos reclutados cumplimentaron una batería de preguntas entre las cuales se encontraban la escala ASES-p (Richards et al. 1994), la escala SF-36 (Ware, Jr. & Sherbourne 1992) y la escala de Barthel (Mahoney & Barthel 1965). La recogida de datos se realizó en el momento basal y a los 3 o 6 meses post intervención, según el tipo de la misma. Los cuestionarios se rellenaron por los propios participantes en su domicilio. Todos los participantes tenían a su disposición el teléfono directo del técnico de investigación contratado para

el seguimiento de este proyecto, disponible durante su jornada laboral, para solucionar posibles dudas de los pacientes.

### ***Fiabilidad test-retest***

La fiabilidad test-retest de la escala ASES-p no se pudo estudiar con los datos recogidos durante este estudio, ya que las respuestas retest no se recibieron en el plazo de tiempo establecido. Por este motivo, dicha fiabilidad se estudió con datos recogidos mediante entrevistas telefónicas, realizadas a posteriori, a los sujetos que proporcionaron respuestas válidas en este estudio. Los respectivos resultados se encuentran en el capítulo 7.

## **4.3 Estudio de evaluación de la escala Constant-Murley Score**

Información más detallada sobre este estudio y los resultados del mismo se presentan en el capítulo 8.

### ***Revisión sistemática y evaluación de la evidencia***

El objetivo de la segunda búsqueda fue encontrar información sobre las propiedades psicométricas de la escala funcional de Constant-Murley Score (CMS), en diferentes patologías de la articulación de hombro. Esta búsqueda se realizó en las bibliotecas PubMed y Embase para el periodo entre 1 de enero 1986 hasta 2 de mayo 2014. No hubo restricciones de idiomas durante la búsqueda, aunque finalmente se consideraron artículos publicados en inglés, español, francés, alemán e italiano. Los artículos finalmente seleccionados se separaron en grupos de diferentes patologías de hombro. Dos evaluadores independientes evaluaron las cualidades psicométricas de CMS en cada grupo, mediante la herramienta EMPRO.

#### 4.4 Herramientas y escalas utilizadas en los diferentes estudios

Esta sección presenta las herramientas y escalas implementadas en los diferentes estudios de esta tesis. Concretamente se dan detalles sobre: las herramienta de evaluación EMPRO y las escalas ASES-p, SF-36, Barthel y CMS. Todas ellas se incluyen además en el Anexo I.

##### **EMPRO**

Como ya se menciona anteriormente, las evaluaciones estandarizadas de los estudios que se presentan en los capítulos 5 y 7 de esta tesis se realizaron mediante la herramienta EMPRO. A continuación se explica de manera detallada cómo se derivan las puntuaciones de esta herramienta.

Como ya se menciona en el capítulo anterior, el EMPRO se compone de 39 ítems que representan los 8 atributos: *concepto y modelo de medida*, *fiabilidad*, *validez*, *sensibilidad al cambio*, *interpretabilidad*, *facilidad de uso del instrumento*, *formas alternativas de administración*, y *proceso de adaptación cultural y lingüística* (Valderas et al. 2008). Cada atributo consiste de 2 a 8 ítems, mientras que cada ítem cuenta con 4 alternativas de respuesta con valores que oscilan desde 1="fuertemente en desacuerdo" hasta 4="fuertemente en acuerdo". Todos los ítems tienen una opción de respuesta de "no información", mientras que algunos ítems cuentan también con la opción adicional "no aplicable". La puntuación de cada atributo se calcula cuando por los menos el 50% de sus respectivos ítems hayan sido respondidos. En estos casos se calcula el valor medio de las respuestas, después de asignar el valor 1 (i.e. mínimo valor posible) a los ítems con respuestas "no información". Cuando los ítems respondidos no llegan al 50%, entonces se considera que el atributo no se puede valorar y no se proporciona puntuación para el mismo.

El atributo de la *fiabilidad*, se divide en dos sub-atributos, la *consistencia interna* y la *reproducibilidad*. La puntuación final asignada a este atributo es la puntuación más alta obtenida de los dos sub-atributos. El atributo de la *facilidad de uso* también se divide en

dos, la carga de uso para los entrevistados y la carga de administración, aunque en este caso las respectivas puntuaciones se presentan de manera individual. Tanto la *facilidad de uso*, como las *formas alternativas de administración* no se incorporan en la puntuación global de EMPRO y se estudian aparte. Cuando se evalúan herramientas originales (i.e. no las adaptadas a otros idiomas) tampoco se considera el atributo del proceso de adaptación cultural y lingüística. La puntuación global de EMPRO se basa en los primeros 5 atributos arriba mencionados y se calcula solo si existe información en por lo menos 3 de ellos. Los atributos sin información se puntúan con 0. Las puntuaciones finales que proporciona EMPRO, tanto la global como las puntuaciones por atributos, oscilan entre un mínimo de 0 hasta un máximo de 100 puntos. Valores aceptables se consideran los  $\geq 50$  puntos (Garin et al. 2013).

Los ítems de EMPRO se acompañan por instrucciones precisas acerca del tipo de información que se debería evaluar en cada caso, así como explicaciones sobre qué puntuación se debería asignar según la información disponible. Adicionalmente, todos los ítems cuentan con espacio para comentarios, mientras que el evaluador tiene que apuntar en qué artículo y página del mismo, se encuentra la información que está puntuando. El EMPRO requiere la participación de evaluadores expertos en el ámbito de las cualidades psicométricas de los cuestionarios. Su formato facilita la comparación y consenso entre las puntuaciones asignadas de los diferentes evaluadores, ya que fue creado precisamente para proporcionar evaluaciones consensuadas.

Las puntuaciones EMPRO, utilizadas en los estudios 4.1 y 4.3, se basaron en evaluaciones de expertos por pares, realizando los siguientes pasos. Primero, todos los evaluadores realizaron de modo independiente la correspondiente evaluación de la evidencia disponible. A continuación, las dos evaluaciones se compararon y se identificaron las discrepancias. Estas últimas se resolvieron con discusión entre las dos partes hasta llegar a una evaluación conjunta y consensuada. Un tercer evaluador estaba disponible para ayudar en los casos de discrepancias difíciles de solucionar. Ninguno de los revisores participantes fue autor de los artículos evaluados. La mayoría de ellos pertenecían al grupo que desarrolló EMPRO y el resto tenía experiencia en la

aplicación de la herramienta, habiendo realizado un curso de formación sobre su aplicación y haberlo utilizado anteriormente.

### **ASES-p**

Como ya se ha expuesto en apartados anteriores, este cuestionario fue el finalmente seleccionado entre todos los evaluados (capítulo 5) y posteriormente adaptado y validado al castellano (capítulos 6 y 7).

El *American Shoulder and Elbow Surgeons Form* (ASES) se publicó en 1994 por la asociación de los cirujanos Americanos de hombro y codo (Richards et al. 1994). La versión extensa de esta escala cuenta con una parte que se rellena por el especialista y otra, auto-cumplimentada por el propio paciente. En la presente tesis la parte auto-cumplimentada se denomina ASES-p (Figura 4.1). Los ítems que corresponden al especialista se relacionan con el rango de movimiento que puede realizar el sujeto con el lado afectado, como por ejemplo la rotación interna o externa, una serie de síntomas relacionados con la patología de hombro y la inestabilidad de la articulación. Las preguntas auto-cumplimentadas recogen información sobre el nivel de dolor que experimenta el paciente en el momento de la cumplimentación (1 ítem) y la dificultad que tiene en realizar una serie de 10 actividades de vida diaria (AVD) que los autores originales engloban bajo la denominación *función*. El *dolor* y la *función* representan las dos dimensiones de la escala, cada una de las cuales cuenta con su propia puntuación.

El ítem de *dolor* se mide con una escala visual analógica (EVA) de 10 cm. Se asignan 0 puntos a la ausencia de dolor y 10 puntos al dolor de máxima intensidad. Los ítems de la *función* cuentan con respuestas tipo Likert con 4 alternativas de respuesta que oscilan entre 0 (imposible) hasta 3 (sin ninguna dificultad). La puntuación de cada dimensión oscila de un mínimo de 0 hasta un máximo de 50 puntos. Para la sub-escala de dolor dicha puntuación es:

$$(10\text{-EVA}) \times 5$$

Para la puntuación de la *función*, se suman las respuestas de todos los ítems de las AVD y se transforman con la siguiente fórmula:

$$\text{Suma de 10 AVD ítems} \times (5/3)$$

La puntuación total de la escala es la suma de ambas puntuaciones y su posible rango oscila entre 0 hasta 100 puntos. Puntuaciones más elevadas representan un mejor estado de la articulación del hombro. El ASES-p se utiliza como escala individual, independientemente del resto de ítems de la versión extensa.

Patient Self-Evaluation												
<b>Pain:</b>												
How bad is your pain today (mark line)?												
0									10			
No pain at all	<div style="display: flex; justify-content: space-between; border-top: 1px solid black; border-bottom: 1px solid black;"> <span style="width: 10%;"></span> <span style="width: 10%;"></span> <span style="width: 10%;"></span> <span style="width: 10%;"></span> <span style="width: 10%;"></span> <span style="width: 10%;"></span> <span style="width: 10%;"></span> <span style="width: 10%;"></span> <span style="width: 10%;"></span> </div>								Pain as bad as it can be			
<b>Function:</b> Circle the number in the box that indicates your ability to do the following activities:												
0 = Unable to do;      1= Very difficult;      2=Somewhat difficult;      3=Not difficult												
Activity	Right arm				Left arm							
1. Put on a coat	0	1	2	3	0	1	2	3	0	1	2	3
2. Sleep on your painful of affected side	0	1	2	3	0	1	2	3	0	1	2	3
3. Wash back/do up bra	0	1	2	3	0	1	2	3	0	1	2	3
4. Manage toileting	0	1	2	3	0	1	2	3	0	1	2	3
5. Comb hair	0	1	2	3	0	1	2	3	0	1	2	3
6. Reach a high shelf	0	1	2	3	0	1	2	3	0	1	2	3
7. Lift 10 lbs above shoulder	0	1	2	3	0	1	2	3	0	1	2	3
8. Throw a ball overhand	0	1	2	3	0	1	2	3	0	1	2	3
9. Do usual work- List: .....	0	1	2	3	0	1	2	3	0	1	2	3
10. Do usual sport- List: .....	0	1	2	3	0	1	2	3	0	1	2	3

**Figura 4.1:** Versión original de la parte auto-administrada del cuestionario American Shoulder and Elbow Surgeons Shoulder Assessment (ASES-p).

**SF-36**

El 36-item Short Form Health Survey (SF-36) es un instrumento auto-administrado, compuesto por 36 preguntas, que explora 8 dimensiones: *función física, rol físico, dolor, salud general, vitalidad, función social, rol emocional y salud mental* (Ware, Jr. & Sherbourne 1992). Dicho cuestionario ofrece una puntuación para cada dimensión y dos puntuaciones resumen, una física (PCS) y otra mental (MCS). Todas las puntuaciones proporcionadas oscilan de un mínimo de 0 puntos (peor estado) hasta un máximo de 100 (mejor estado). Siendo una escala genérica se puede usar tanto en población sana o pacientes de cualquier patología (Brazier et al. 1992, Alonso et al. 1998). Es uno de los

cuestionarios más referenciados en la literatura científica, especialmente en estudios de CVRS y la evidencia disponible justifica su extendido uso (Vilagut et al. 2005). Aparte de la versión extensa, también existen dos versiones reducidas de 12 y 8 ítems, las denominadas SF-12 y SF-8 (Ware, Jr. et al. 1996, Ware, Jr. et al.). En el presente trabajo, se administró la versión SF-36 en español, obtenida mediante licencia por la empresa QualityMetric Incorporated.

### **CMS**

El Constant-Murley Score (CMS), como ya se expone en otras secciones de esta tesis, es un cuestionario creado para la evaluación del hombro afectado, que combina la observación clínica (datos *objetivos*) con preguntas auto-cumplimentadas por el propio paciente (datos *subjetivos*). La parte subjetiva recoge información sobre el dolor y las actividades de la vida diaria (AVD), mediante 2 y 4 ítems respectivamente. Mientras que la parte objetiva tiene que ver con, el rango de movimiento y la fuerza de la articulación que se evalúa por 4 y 1 ítem. Cada uno de componentes arriba mencionado recibe un máximo de 15, 20, 40 y 25 puntos respectivamente; con puntuación total de la escala la suma de estos componentes. Una puntuación de 100 puntos representaría un hombro en estado óptimo, mientras que 0 puntos implicarían una articulación en pésimo estado. El 65% de su puntuación total de este instrumento se basa en los datos de observación I clínico (Constant & Murley 1987).

A lo largo de los años el CMS se ha implementado en un gran número de proyectos relacionados con la articulación de hombro y se considera por muchos, con la aprobación y apoyo de la Sociedad Europea de Cirugía de Hombro y Codo (SECEC), como la regla de oro en la valoración de esta articulación (Barra-Lopez 2007, Lillkrona 2008). Cabe mencionar que la primera publicación sobre esta herramienta evaluaba el dolor con una escala Likert de 4 alternativas de respuesta, mientras que en una publicación posterior, entre otros cambios, el autor original aconsejaba la evaluación del dolor con una escala visual analógica de 0-15 puntos (Constant et al. 2008). La versión de CMS proporcionada en la página web de SECEC incluye la evaluación del dolor de ambas maneras, con puntuación final la media de las dos puntuaciones (European Society for



Surgery of the Shoulder and the Elbow 2016). Esta es la versión que se implementó en el proyecto 4.2 de la presente tesis. Además, aparte de la puntuación original de CMS (CMS<sub>O</sub>), siguiendo las últimas indicaciones proporcionadas por su autor, también se calculó una puntuación de CMS ajustada por la edad y el sexo (CMS<sub>R</sub>) de los participantes (Constant et al. 2008). Para obtener dicho ajuste, las puntuaciones obtenidas por el CMS original tienen que dividirse con los valores medios “normales”, que los sujetos deberían presentar según su edad y sexo (Figura 4.2). Estos últimos valores están basados en un estudio de 900 sujetos sanos llevado a cabo por el mismo autor (Constant 1986). De manera adicional, se exploró una versión de CMS sin el componente de la fuerza (CMS<sub>NS</sub>), como sugieren otros autores (Othman & Taylor 2004), mientras que los cuatro componentes de CMS también se estudiaron individualmente.

**Table II** Constant score normal values for different age and sex groups

Age	Male subjects				Female subjects			
	Right	Left	Mean	SD	Right	Left	Mean	SD
21-30 y	97	99	98	4.2	98	96	97	4.7
31-40 y	97	90	93	3.4	90	91	90	4.1
41-50 y	86	96	92	3.6*	85	78	80	3.8*
51-60 y	94	87	90	3.1*	75	71	73	2.8*
61-70 y	83	83	83	4.2*	70	68	70	4.0*
71-80 y	76	73	75	3.6	71	64	69	3.9
81-90 y	70	61	66	3.1	65	64	64	2.9
91-100 y	60	54	56	4.3	58	50	52	5.1

Data are from Constant.<sup>6</sup>  
\*Significant difference compared with opposite sex.

**Figura 4.2:** Valores CMS de sujetos sanos, en diferentes grupos de edad y sexo. La tabla se ha copiado del artículo Constant et al. 2008.

### **Barthel**

El índice de Barthel es un cuestionario diseñado para evaluar la capacidad de los individuos en realizar actividades básicas de la vida diaria (ABVD). Se creó para monitorizar los cambios producidos en pacientes hospitalizados, después de recibir una intervención (Mahoney et al. 1958). En la literatura existen varias versiones de este cuestionario. En este proyecto se administró una de la más implementada, la versión de 10 ítems que pregunta sobre cuidado personal y movilidad (Mahoney & Barthel 1965). Los ítems de esta escala tienen entre 2 y 4 alternativas de respuesta y el valor máximo de cada ítem se mueve entre 5 y 15 puntos. La puntuación total es la suma de las puntuaciones de los 10 ítems y su rango es de 0 a 100, con intervalos de 5 puntos. El valor 100 indica que el sujeto es independiente en cuanto a sus ABVD, el valor 95 indica que el sujeto necesita algún tipo de ayuda, en una de las actividades exploradas, mientras que valores de 90 o menos puntos sugieren obviamente estados con diferentes grados de dependencia.

## **4.5 Análisis estadístico**

En esta sección se exponen los métodos estadísticos utilizados en la validación de la escala ASES-p, cuyos resultados se presentan en los capítulos 6 y 7 de la presente tesis.

La validación de una escala requiere la aplicación de varios métodos estadísticos. Entre ellos, se encuentran pruebas relativamente sencillas, como pueden ser por ejemplo los coeficientes de correlación, para comprobar validez convergente o discriminante entre los ítems de la escala o con otras medidas, el t-test de Student para comprobar la significación de las diferencias observadas en las puntuaciones obtenidas entre dos momentos o los efectos estandarizados para evaluar la sensibilidad al cambio. Pero también existen métodos más avanzados, como son los análisis factoriales, tanto exploratorios (AFE) como confirmatorio (AFC) y los modelos de la respuesta al ítem, entre los cuales se encuentra el modelo Rasch. Estos métodos se implementan para

estudiar la validez de constructo de una escala, ya que proporcionan información sobre las relaciones entre los diferentes ítems de las mismas, y sobre los factores latentes (no observados) que se supone que explican dichas relaciones.

#### 4.5.1 Fiabilidad

##### ***Consistencia Interna***

La consistencia interna entre los ítems de la escala ASES-p se evaluó con el coeficiente alpha de Cronbach (Bland & Altman 1997). Este coeficiente es una función del número de los ítems de la escala (k), la varianza de cada ítem ( $s_i^2$ ) y la varianza de la puntuación total, derivada de la suma de todos los ítems ( $s_T^2$ ). El alpha de Cronbach se estima de la siguiente fórmula:

$$a = \frac{k}{k-1} \left(1 - \frac{\sum s_i^2}{s_T^2}\right)$$

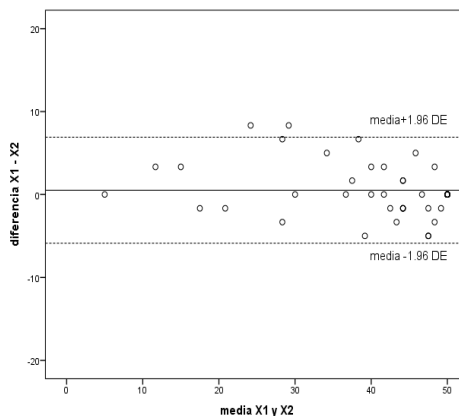
Sus posibles valores oscilan entre 0 y 1, representando una escala con ítems completamente independientes y una escala con ítems idénticos, respectivamente. Aparte de este coeficiente, también se exploraron las correlaciones entre los diferentes ítems (ítem-ítem) y entre cada ítem y la puntuación total de la escala (ítem-total), tras corregir por solapamiento, es decir, después de eliminar el correspondiente ítem de la puntuación total. Se utilizó la correlación de Spearman ( $r_s$ ), que se basa en el orden de las observaciones. Se consideraron aceptables valores alpha de Cronbach  $\geq 0.70$  y valores  $r_s \geq 0.30$  (Terwee et al. 2007).

##### ***Fiabilidad test-retest***

La fiabilidad test-retest del cuestionario ASES-p se basó en las respuestas de dos administraciones consecutivas, realizadas con una diferencia de 3-7 días entre ellas. Dicha fiabilidad se comprobó de distintas maneras. Por un lado, se estudió la concordancia exacta, es decir el porcentaje de diferencias igual a 0, en todos los ítems y

las puntuaciones de la escala. También se estimó el coeficiente de correlación Intraclase de las puntuaciones totales mediante un modelo mixto, donde los sujetos eran el efecto aleatorio y los evaluadores los efectos fijos (CCI:3,1) (Shrout & Fleiss 1979).

Por último se calcularon los límites de concordancia al 95% (Limits of agreement), con el método propuesto por Bland y Altman (Bland & Altman 1986). Los límites de concordancia evalúan las diferencias entre mediciones consecutivas (en diferentes momentos o por diferentes observadores) frente al valor real de las mismas. Dado que el valor real no se conoce, como tal se considera la media de las mediciones. Los límites de concordancia son una función de la media de las diferencias ( $\bar{d}$ ) y la desviación estándar (DE) de las mismas. Dado que las diferencias entre mediciones consecutivas suelen tener una distribución Normal, se considera que el 95% de ellas debería encontrarse entre  $\bar{d} - 1.96 DE$  hasta  $\bar{d} + 1.96 DE$  (Figura 4.3). Si este rango de diferencias no es clínicamente importante, entonces se considera que existe concordancia entre las mediciones. El valor  $1.96 DE$  se denomina coeficiente de fiabilidad. En los últimos años se ha observado una tendencia creciente en la aplicación de este método (Dewitte et al. 2002).



**Figura 4.3:** Grafico de dispersión de Bland y Altman. La línea continua representa la media diferencia entre dos métodos y las líneas discontinuas los respectivos límites de concordancia al 95%.

### 4.5.2 Validez

Existen diferentes aproximaciones matemáticas para estudiar la validez de constructo de una escala. Los modelos de análisis factorial (modelos de la teoría clásica del test) y el modelo Rasch (modelos de la teoría de la respuesta al ítem) se utilizan para comprobar la validez estructural de un cuestionario.

#### ***Validez estructural: análisis factorial***

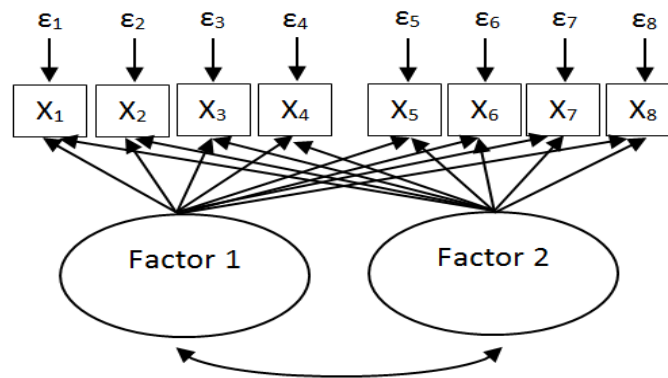
Desde su concepción, hace ya más de un siglo (Spearman 1904) el análisis factorial se ha establecido como uno de los análisis multivariantes más utilizados. El principal objetivo de este tipo de análisis es estudiar e intentar establecer la naturaleza de los *factores latentes* que explican la varianza y covarianza de una serie de *indicadores*. Los *indicadores* son variables medibles y observadas (los ítems), mientras que los *factores latentes* son variables no observables. Estos modelos tienen su origen en la *teoría del factor común*, que postula que cada indicador de un conjunto de indicadores, es una función lineal de uno o más factores comunes entre los indicadores y de un factor único para cada indicador. Por lo tanto, el análisis factorial parte la varianza de cada indicador en dos componentes: *a) la varianza común*, la que explica el factor común y *b) la varianza única* de cada indicador que se debe a factores latentes específicos para cada indicador y a errores aleatorios. Entre otras estimaciones, estos modelos proporcionan coeficientes de cargas (o pesos) factoriales, que son el efecto que cada factor tiene sobre cada indicador. La magnitud de estos pesos indica la importancia de los respectivos ítems en relación al constructo que se estudia (Brown 2006).

Los modelos factoriales se usan frecuentemente en psicología, sociología, educación y ciencias de la salud, para la evaluación de cuestionarios multi-ítem. Se utilizan tanto para crear cuestionarios, como para obtener evidencia sobre su validez de estructura, o incluso para reducir el número de ítems de una escala (Brown et al. 2005, Bonner et al. 2016). Existen dos principales tipos de análisis basados en la exploración del factor común: los modelos de análisis factorial exploratorio (AFE) y los modelos de análisis

factorial confirmatorio (AFC). Ambos tipos de modelos estudian la relación que un conjunto de ítems tiene con un número reducido de factores latentes. Su principal diferencia se encuentra en las especificaciones y restricciones que cada uno de estos modelos impone a los componentes arriba expuestos. Aunque para las necesidades de la presente tesis se realizaron solo modelos confirmatorios, a continuación exponemos brevemente ambos tipos de modelos.

### **Análisis factorial exploratorio**

Los modelos de AFE se suelen implementar en el proceso inicial del desarrollo de un cuestionario. En esta fase, los investigadores estudian el comportamiento de los ítems de interés y comprueban diferentes hipótesis, sobre los factores latentes que pretenden estudiar. Aunque se tiene que determinar el número de factores y la relación entre ellos, es decir si están correlacionados o no, la relación de los ítems con los diferentes factores se establece de manera “libre” ya que dicha relación, se determina por los datos y las estimaciones obtenidas, no se especifica a-priori. Un supuesto modelo de AFE de dos factores y ocho ítems se presenta la Figura 4.4.

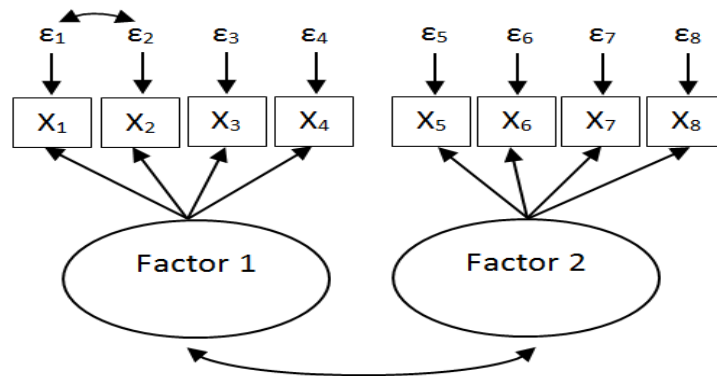


**Figura 4.4:** Modelo de análisis factorial exploratorio con dos factores correlacionados.  $X_i$ : representa los ítems observados,  $\epsilon_i$ : representa los errores aleatorios de cada ítem.

Las cargas factoriales estimadas por este modelo indicarán la relación entre factores e ítems, y guiarán la asignación de cada ítem a su correspondiente factor. En los modelos de AFE los errores aleatorios se consideran independientes entre ellos. La denominación “exploratorio” se debe al hecho de que estos modelos utilizan los datos disponibles, principalmente para generar hipótesis (Brown 2006, Flora et al. 2012).

### **Análisis factorial confirmatorio**

Por otro lado, los AFC parten de una teoría o una hipótesis previa concreta y su función es comprobar la confirmación de la misma. Los AFC se suelen implementar para comprobar la validez y el funcionamiento de una escala, por ejemplo con datos de otros ámbitos, en validaciones culturales, o con pacientes de otras patologías y características, diferentes para los que se diseñó originalmente la escala. La aplicación de estos modelos requiere conocimiento previo sobre el número de factores latentes que tiene la escala, los ítems que se asocian con cada factor, así como las relaciones entre ítems (si existe correlación entre algunos de ellos) y entre los factores (si están correlacionados o no). Para realizar un modelo de AFC, es importante disponer de información previa, reflexionar y tomar decisiones sobre todos estos aspectos (Brown 2006, Flora et al. 2012). Un supuesto modelo de AFC se presenta en la Figura 4.5.



**Figura 4.5:** Modelo de análisis factorial confirmatorio con dos factores relacionados y correlación entre dos errores. X<sub>i</sub>: representa los ítems observados,  $\epsilon_i$ : representa los errores aleatorios de cada ítem.

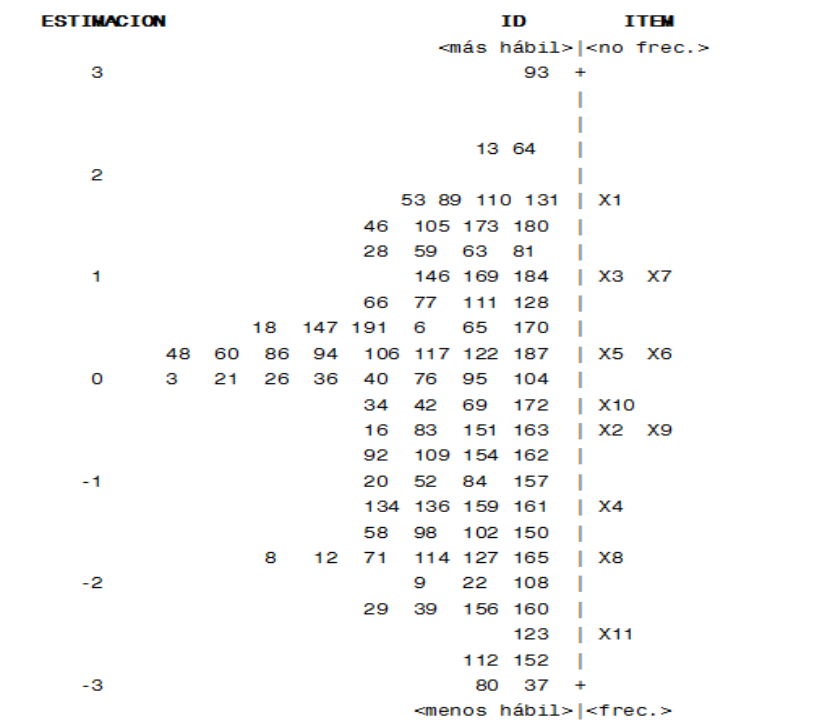
Los modelos de AFC que se realizaron como parte del proceso de validación del cuestionario ASES-p, fueron de un único factor. Aunque el ASES-p se compone de dos sub-escalas (dolor y función), el hecho que la sub-escala del dolor se basa en una única pregunta, no permitía realizar un modelo de dos factores, uno de ellos con único indicador el ítem del dolor, ya que este modelo no podía converger, ni producir estimaciones (Brown 2006). Por este motivo se estudiaron modelos de un factor con y sin el ítem del dolor. Se implementó el método de estimación de *Unweighted Least Squares* (ULSMV), recomendada para indicadores ordinales o no-normales y muestras  $n < 200$  sujetos (Forero et al. 2009). Los índices de ajuste estudiados y sus correspondientes valores aceptables fueron los siguientes: Root Mean Square Error of Approximation (RMSEA)  $< 0.08$ ; Tucker-Lewis Index (TLI)  $> 0.90$  y Comparative Fit Index (CFI)  $> 0.90$ . Se consideraron aceptables pesos factoriales  $\geq 0.40$ . Finalmente se estudiaron los valores residuales y los índices de modificación que pueden sugerir ciertas modificaciones para mejorar el ajuste del modelo (Brown 2006).

#### **Validez estructural: el modelo Rasch**

La validez de constructo de una escala también se puede estudiar con el modelo Rasch. Este modelo se inventó por el matemático Danés Georg Rasch (1901-1980) y pertenece a los modelos de la teoría de respuesta al ítem (TRI). El modelo Rasch estudia la unidimensionalidad de un cuestionario. Por este motivo, en caso de instrumentos con varias dimensiones, este modelo se suele aplicar por separado a cada una de ellas. Este modelo se puede usar tanto para crear una escala, como para confirmar su estructura o para reducir el número de sus ítems. El modelo Rasch, incorpora el concepto del orden, tanto para los ítems de la escala (*dificultad*) como para los sujetos (*habilidad*). Consideremos, por ejemplo, una escala que evalúa la capacidad en realizar ciertas AVD con una serie de ítems con respuestas binarias (Si/No). En este caso, el ítem más *difícil* sería aquel al que todos los sujetos contestan “No”, mientras que el más fácil el que todos los sujetos contestan con un “Sí”. De manera similar los sujetos más y menos *hábiles* serían los que responden Si y No a todos los ítems de la escala respectivamente. El modelo Rasch suma las respuestas recibidas en cada ítem, calcula los porcentajes de Sí y No de todos los ítems, estima el ratio de los mismos y, a continuación, transforma este ratio a su logaritmo natural. De esta manera, las puntuaciones crudas pasan a una escala logit que representa mejor las distancias relativas entre ellas. El



mismo proceso se realiza en el caso de los sujetos. Los sujetos con mayor habilidad son más probables en contestar positivamente a los ítems de mayor dificultad. Un ejemplo gráfico de las estimaciones producidas por un modelo Rasch, se presenta en la Figura 4.6. En ella se aprecia que el sujeto más hábil es el id=93, mientras que el ítem más difícil es el X1. En el otro extremo de la escala, los sujetos menos hábiles son los id=37, 80 y el ítem más fácil el X11.



**Figura 4.6:** Ejemplo de mapa producido por un modelo Rasch. La habilidad de las personas (ID) y la dificultad de los ítems se estudia en la misma escala logit.

El modelo Rasch se puede aplicar a cuestionarios con ítems de respuestas politómicas, tipo Likert. Esta extensión se puede realizar con el modelo Andrich Rating Scale Model (Bond & Fox 2007, Wright B.D 1998, Pallant & Tennant 2007), que se utilizó en la validación de ASES-p. Para poder introducir el ítem del dolor en este modelo, las respuestas recibidas en el mismo se transformaron en una variable ordinal de 4 categorías. Se estimaron los valores logit de la *dificultad* de los ítems. La bondad de ajuste se comprobó de las siguientes

maneras. Por un lado se estimaron los índices de Infit (*inlier-pattern-sensitive fit statistic*) y Outfit (*outlier-sensitive fit statistic*) Mean Square (MNSQ). Los valores aceptables de estos índices se tienen que situar entre 0.6 y 1.4, mientras que valores  $>2$  implican que los ítems distorsionan las mediciones de la escala, por lo que se aconseja su eliminación de la misma (Wright & Linacre 1994, Bond & Fox 2007).

Se estudiaron los estadísticos de *fiabilidad* y de *separación* de los sujetos y de los ítems. El índice de *fiabilidad de los sujetos* (*person reliability*) indica como de reproducible es el orden de la habilidad de los sujetos, para el conjunto de los ítems que se está estudiando. La *fiabilidad de los ítems* (*items reliability*) indica como de reproducible es el orden de dificultad de los ítems, en esta muestra de sujetos. En ambos casos, los valores deseables deberían ser  $>0.80$ . El índice de *separación de los sujetos* (*person separation*) evalúa la capacidad de la escala en distinguir entre diferentes grupos de habilidad o de gravedad, de los participantes, mientras que el estadístico de la *separación de los ítems* (*item separation*) evalúa la jerarquía de la dificultad de los ítems. Valores  $>2$  se esperan en ambos índices. Finalmente, la unidimensionalidad de la escala se comprobó mediante el análisis de componentes principales realizados con los residuos del modelo Rasch. Valores eigenvalue  $<3$  apoyan la unidimensionalidad del instrumento, ya que indican la inexistencia de factores latentes adicionales (Linacre 2011).

### ***Validez convergente y discriminante***

Se estudió la validez convergente y discriminante de ASES-p con las escalas de CMS, SF-36 y Barthel, mediante la correlación de Spearman. Correlaciones más elevadas con las escalas CMS, Barthel, y con las dimensiones y el componente físico de SF-36 apoyarían la validez convergente de ASES-p. Por otro lado, correlaciones más bajas con las dimensiones y componentes mentales de SF-36 apoyarían la validez discriminante. Asimismo, las dos sub-escalas de ASES-p deberían presentar correlaciones más elevadas con los respectivos componentes de CMS y dimensiones de SF-36, mientras que, en todo caso, las correlaciones con los componentes y dimensión mental de SF-36 deberían ser más bajas.

### **Validez de grupos conocidos**

Para estudiar la validez de grupos conocidos, se categorizaron las escalas CMS<sub>0</sub>, PCS de SF-36 y Barthel en diferentes grupos de gravedad, y se compararon los valores de ASES-p observados en estos grupos.

Las puntuaciones CMS<sub>0</sub> y Barthel se dividieron en tres grupos. Concretamente, para el CMS<sub>0</sub> se consideraron los valores  $\leq 30$ , 31-60, y  $\geq 61$  puntos, mientras que los grupos de Barthel fueron 80-90, 95 y 100 puntos. La categorización de CMS<sub>0</sub> fue arbitraria, debido a la falta de consenso acerca de los puntos de corte de esta escala. Por otro lado, la categorización de Barthel correspondía a sujetos completamente autónomos (100 puntos); sujetos con algún problema en una de las ABVD (95 puntos); y finalmente sujetos con 2 o más problemas (80-90 puntos) (Cid-Ruzafa & Damian-Moreno 1997). Debido a que los valores ASES-p, obtenidos en estos grupos, no tenía una distribución Normal, la respectiva información se estudió mediante los valores de las medianas. Se calcularon las medianas y los rangos intercuartílicos (Q1, Q3), de las puntuaciones ASES-p, en los diferentes grupos. La comparación de estos valores se realizó con el test de Jonckheere-Terpstra. Se trata de una prueba no paramétrica que se utiliza para comparar los valores observados en 3 o más grupos ordinales y comprobar la tendencia de aumento (o disminución), de dichos valores, entre los consecutivos grupos (Walker 2002).

En el caso del PCS SF-36 se derivaron dos grupos de puntuaciones. Se consideró que los sujetos con PCS  $< 50$  puntos tenían peor estado de salud, mientras que los sujetos con  $\geq 50$  puntos disfrutaban de un mejor estado de salud (Vergara et al. 2012). En este caso, dado que la distribución de los valores ASES-p fue Normal, la respectiva información se presenta con medias (DE), mientras que las comparaciones entre dos grupos se realizaron con la prueba t-test de Student para muestras independientes. La validez de grupos conocidos se comprobaría si las puntuaciones de ASES-p incrementaran según el grupo de gravedad, establecido por las otras tres escalas.

### 4.5.3 Sensibilidad al cambio

La sensibilidad al cambio se estudio comparando los efectos estandarizados en dos grupos de mejoría (Sí/No) y en tres grupos de tratamientos (intervención quirúrgica, infiltración y otro tratamiento). La mejoría se basó en dos preguntas de auto-evaluación, sobre el rango de movimiento del hombro afectado y la capacidad a realizar las AVD.

Con las respuestas pre-post de ASES-p se calcularon tres tipos de efectos: el efecto estandarizado (SES), la respuesta media estandarizada (SRM) y la respuesta media estandarizada ajustada para datos emparejados ( $SRM_{Adj}$ ). Los dos primeros efectos son una función de la media diferencia entre los dos momentos ( $\bar{d}$ ) y la desviación estándar de los valores basales ( $DE_{basal}$ ) o de las diferencias ( $DE_d$ ).

$$SES = \frac{\bar{d}}{DE_{basal}}$$

$$SRM = \frac{\bar{d}}{DE_d}$$

El  $SRM_{Adj}$  es una función de un efecto basado en la combinación de las desviaciones de los valores basal y final ( $ESp$ ) y de la correlación ( $r$ ) entre las dos medidas. Se considera que este efecto puede proporcionar estimaciones de efectos más adecuadas, en los casos de medidas emparejadas (Middel & van 2002).

$$ESp = \frac{\bar{d}}{\sqrt{(DE_{basal}^2 + DE_{final}^2)/2}}$$

$$SRM_{Adj} = \frac{ESp/\sqrt{2}}{\sqrt{(1-r)}}$$

Según la definición de Cohen, efectos de 0.20, 0.50 y 0.80 se describirían como pequeños, medianos y grandes respectivamente. Para las necesidades de este trabajo se consideró que ciertos grupos de pacientes deberían presentar efectos más elevados, en comparación con los demás. Concretamente, aquellos pacientes cuya auto-evaluación del estado del hombro afectado había mejorado, y los pacientes que se sometieron a una intervención quirúrgica. En relación a estos últimos, se supuso que los sujetos intervenidos quirúrgicamente son,

generalmente, los que peor estado de salud tienen. Por este motivo, el beneficio experimentado después de la intervención y el periodo de la recuperación, debería ser más elevado en este grupo. Consecuentemente, los efectos de este grupo también deberían ser más elevados que los efectos de los otros tratamientos.

#### **4.5.4 Software**

Los modelos de AFC se realizaron con el software MPlus versión 7.4. El modelo Rasch se realizó con Winsteps 3.91.0. Las puntuaciones EMPRO se estimaron con el software SPSS. Las puntuaciones de SF-36 se realizaron con el QualityMetric Health Outcomes™ Scoring Software 4.5. El resto de los análisis se realizó con el software SAS for Windows.



## **5. Escalas auto-cumplimentadas para la patología de hombro: revisión sistemática y evaluación estandarizada de la evidencia disponible**

Los datos presentados en este capítulo se han publicado en la revista "Journal of Shoulder and Elbow Surgery": Evaluation of shoulder-specific patient-reported outcome measures: a systematic and standardized comparison of available evidence. J Shoulder Elbow Surg. 2014 Mar; 23(3):434-44. Schmidt S, Ferrer M, González M, González N, Valderas JM, Alonso J, Escobar A, Vrotsou K.

## 5.1 Introducción

Como ya se menciona en la introducción de esta tesis doctoral, las patologías del hombro, debido sobre todos a la complejidad de su estructura anatómica, suelen ser múltiples y variadas. Aparte de los gastos médicos que generan, también pueden afectar la calidad de vida, las actividades sociales y hasta la estabilidad emocional de los pacientes (Virta et al. 2012, Paananen et al. 2011, van der Heijden 1999). Tradicionalmente, la evaluación de la patología de hombro se hacía enfocando en la capacidad funcional de la articulación, estudiando especialmente el rango de movimientos, la fuerza y el dolor del lado afectado (Constant & Murley 1987). En los últimos años, esta práctica está cambiando, debido sobre todo a los avances del concepto de la calidad de vida relacionada con la salud (CVRS) y la importancia que está adquiriendo en el ámbito clínico la percepción que el propio paciente tiene sobre su estado de salud (Black 2013). En la literatura científica existen varios cuestionarios de CVRS auto-administrados, especialmente diseñados para la evaluación del hombro (Roe et al. 2013). Se han desarrollado escalas para patologías concretas, como es el manguito rotador (Kirkley et al. 2003) o la inestabilidad (Kirkley et al. 1998) y, también escalas para poblaciones con características particulares, como son las personas en silla de ruedas (Curtis et al. 1995). Asimismo otros instrumentos se crearon para ser aplicados a cualquier patología de esta articulación (Roe et al. 2013, Wright & Baumgarten 2010, Beaton & Richards 1996).

En los capítulos anteriores se han expuesto las propiedades deseables de los instrumentos de CVRS. Además la selección del más apropiado a utilizar en cada caso, depende de los objetivos del estudio, las características de la escala y los recursos disponibles (Terwee et al. 2007). Publicaciones previas han intentado evaluar cuestionarios auto-administrados de CVRS para pacientes con patologías de hombro. Sin embargo, entre ellas, algunas no evaluaron todas las características psicométricas pertinentes, mientras que los resultados de ninguno de estos trabajos permitía una comparación directa entre las distintas escalas (Angst et al. 2004, Beaton & Richards 1996, Kirkley et al. 2003, Michener & Leggin 2001, Oh et al. 2009, Placzek et al. 2004, Romeo et al. 1996, Roe et al. 2013, Bot et al. 2004). El artículo de Bot et al. 2004, sí que realizó su evaluación mediante una herramienta basada en las sugerencias de Medical Outcome Trust (Bot et al. 2004). Pero por un lado, el sistema de



puntaje que utilizó (calidad: *buena, dudosa, pobre, no información*) tampoco facilitaba las comparaciones, mientras que debido al tiempo transcurrido la información pertinente tenía que ser actualizada. Por lo tanto la información previamente disponible no nos permitía establecer con seguridad cuál sería la escala de CVRS para patologías de hombro, más apropiada para adaptar y validar al castellano. Para poder realizar dicha selección, el presente estudio implementó la herramienta EMPRO (Evaluating Measures of Patient Reported Outcomes) (Valderas et al. 2008).

El EMPRO fue creado para evaluar las propiedades métricas de escalas auto-cumplimentadas por el paciente. Esta herramienta se desarrolló basada en las recomendaciones del Scientific Advisory Committee del Medical Outcomes Trust (Aaronson et al. 2002). Su evaluación se basa en 8 atributos y un total de 39 ítems, mientras que su administración requiere la participación de expertos en el ámbito de las propiedades psicométricas. Información más detallada sobre EMPRO se puede encontrar en los capítulos 3 y 4. El principal objetivo de este estudio fue la realización de una evaluación sistemática y estandarizada, en base a la evidencia disponible, sobre el proceso de creación, las cualidades psicométricas y de administración de instrumentos de CVRS multi-ítem (i.e. más de tres ítems). De interés fueron las escalas aplicables a todas las patologías de hombro. No se consideraron cuestionarios creados para patologías concretas, o grupos concretos de pacientes.

## **5.2 Material y Métodos**

### ***Búsqueda bibliográfica***

Para detectar los cuestionarios específicos de CVRS sobre patologías de hombro se realizó en marzo 2011 una búsqueda sistemática en la base de datos PubMed. Se establecieron los términos de la búsqueda y se implementaron tanto términos propios del Medical Subject Headings (MeSH), como texto libre.

Se utilizaron los siguientes términos: #1 *Quality of life*, #2 *Questionnaires*, #3 *Cross-Cultural Comparison*, #4 *Disability Evaluation*, #5 *Shoulder*, #6 *Shoulder Pain*, #7 *Shoulder Joint*, #8 *Rotator Cuff* (Schmidt et al. 2014).

Se realizaron búsquedas para cada término por separado y posteriormente una búsqueda múltiple: ((#1) OR (#2) OR (#3) OR (#4)) AND ((#5) OR (#6) OR (#7) OR (#8)).

La revisión de las publicaciones obtenidas se realizó en tres pasos por dos revisoras, una de ellas, la autora de la presente tesis:

1. Se revisaron los títulos de los artículos encontrados.
2. A continuación se revisaron los abstract de los títulos inicialmente incluidos.
3. Por último se revisaron los textos enteros de todas aquellas publicaciones que llegaron a la fase de la evaluación final.

De modo adicional y para cumplimentar la fase de la búsqueda bibliográfica, se revisaron manualmente las listas bibliográficas de todos los artículos finalmente seleccionados. Ambas evaluadoras tenían amplia experiencia en medidas de CVRS y en el ámbito de las cualidades psicométricas. En cada paso, las decisiones tomadas se consensuaron mediante discusión entre las dos partes. Un tercer evaluador se consultó en casos de discrepancias difíciles de resolver.

### ***Criterios de inclusión y exclusión de los artículos***

Se incluyeron aquellos artículos con información sobre: el proceso de creación, cualidades psicométricas o de administración de los cuestionario de CVRS específicos para la patología de hombro. Se excluyeron artículos sobre: instrumentos diseñados para el miembro superior en general (i.e. brazo/codo/hombro) o para pacientes con condiciones músculo-esqueléticas no específicamente diseñados para el hombro. Asimismo se excluyeron cuestionarios diseñados solo para patologías específicas (ej. artrosis, inestabilidad) o grupos concretos (ej. pacientes en silla de ruedas o atletas). También se excluyeron instrumentos que no fueron multi-ítem, considerando como tal herramientas con un mínimo de tres ítems (Brown 2006). Finalmente se excluyeron protocolos de estudios, abstract presentados en congresos, estudios de casos y estudios de animales o cadavéricos. De la evaluación en sí, también se excluyeron las revisiones sistemáticas, aunque las referencias bibliográficas de estas últimas sí que se revisaron. En este proyecto se consideraron solo artículos escritos en inglés.

### **Evaluación estandarizada con EMPRO**

Los artículos finalmente seleccionados se revisaron para concretar el número total de cuestionarios de CVRS utilizados en ellos. A continuación, todos los artículos incluidos se separaron en grupos, según el cuestionario o cuestionarios utilizados en ellos. Un mismo artículo podría entrar en tantos grupos, como cuestionarios implementados en él. El objetivo de cada grupo era evaluar las características psicométricas de cada una de las escalas encontradas en los respectivos artículos. Dos revisores independientes evaluaron los artículos de cada grupo con el instrumento EMPRO. Cada revisor participó en la evaluación de una solo escala. Las evaluaciones se realizaron de manera independientes por cada experto y fueron finalmente consensuadas por cada “pareja” de evaluadores. Los evaluadores tenían experiencia con las escalas auto-cumplimentadas (6 pertenecían al grupo que desarrollo EMPRO y el resto 16 se habían acreditado en un curso previo). Ninguno de ellos fue autor, ni participó en el desarrollo de los instrumentos evaluados.

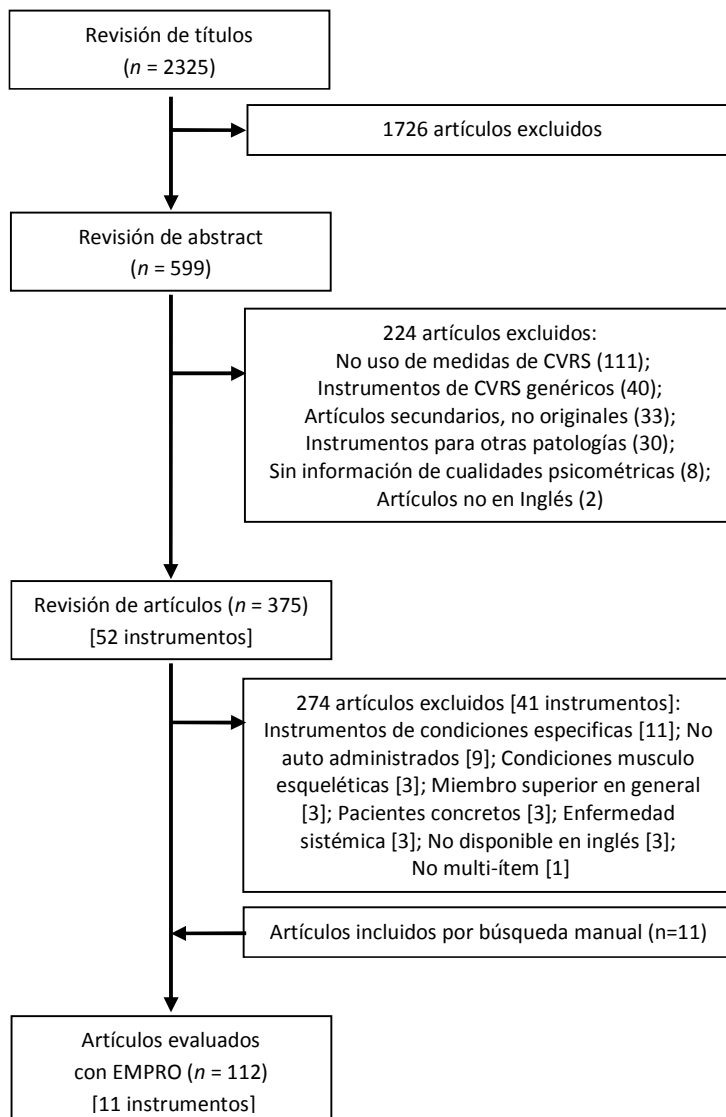
Los atributos EMPRO evaluados en este proyecto fueron los siguientes: *concepto y modelo de medida, fiabilidad (consistencia interna y reproducibilidad), validez, sensibilidad al cambio, e interpretabilidad*. También se evaluaron la *facilidad de uso (para el entrevistado y carga de administración)*, así como las *formas alternativas de administración*, aunque estos últimos atributos no se incluyen en la puntuación total de EMPRO. Las puntuaciones totales y por atributo se estimaron según lo indicado en la sección 4.4. Se consideraron aceptables puntuaciones de 50 o más puntos (Garin et al. 2013).

## **5.3 Resultados**

### **Búsqueda bibliográfica**

En total, en la búsqueda sistemática, se identificaron 2.325 artículos (Figura 5.1). Durante la revisión de títulos, 1.726 artículos se excluyeron por no estar relacionados con el tema del presente proyecto. Otros 224 artículos se excluyeron en la fase de la revisión de los abstract por los siguientes motivos: no contener información sobre medidas de CVRS (111); proporcionar información solo sobre instrumentos genéricos (40); por ser artículos no

originales, de revisión, cartas de opinión u abstract de conferencias (33); por hacer uso de instrumentos de CVRS para patologías distintas al hombro (30); artículos sin ninguna información sobre propiedades psicométricas(8) y artículos no disponibles en inglés (2). Por lo tanto, en total se revisaron los textos de 375 artículos con información sobre 52 instrumentos para la evaluación de la patología de hombro.



**Figura 5.1:** Diagrama de flujo con información sobre la revisión sistemática. Se presenta información sobre el número de artículos y número de instrumentos incluidos y excluidos.

De ellos, 41 instrumentos (274 artículos) se excluyeron por: estar relacionados con medidas de CVRS de patologías específicas de hombro; por estudiar patologías distintas a la patología de interés; por no ser auto-administrados ni multi-ítem; cuestionarios diseñados para grupos específicos; instrumentos sobre condiciones musculo-esqueléticas o miembro superior en general; enfermedades sistémicas y finalmente cuestionarios no disponibles en inglés. Después de revisar manualmente las listas bibliográficas de los artículos identificados se identificaron 11 artículos adicionales que también entraron en la fase de la evaluación final. Así que en la evaluación final se incluyeron un total de 112 artículos con información sobre las cualidades psicométricas de 11 cuestionarios.

Los cuestionarios finalmente evaluados fueron los siguientes:

1. **SSI**: Shoulder Severity Index (Patte D. 1987)
2. **SPADI**: Shoulder Pain and Disability Index (Roach et al. 1991)
3. **SST**: Simple Shoulder Test (Lippitt et al. 1993)
4. **SDQ-UK**: United Kingdom Shoulder Disability Questionnaire (Croft et al. 1994)
5. **ASES-p**: American Shoulder and Elbow Surgeons shoulder assessment – patient self-evaluation section (Richards et al. 1994)
6. **OSS**: Oxford Shoulder Score (Dawson et al. 1996)
7. **SRQ**: Shoulder Rating Questionnaire(L'Insalata et al. 1997)
8. **SSRS**: Subjective Shoulder Rating System (Kohn & Geyer 1997)
9. **PSS**: Penn Shoulder Score (Leggin & Lannotti 1999)
10. **SDQ-NL**: Dutch Shoulder Disability Questionnaire (van der Heijden et al. 2000)
11. **FLEX-SF**: Flexilevel Scale of Shoulder Function (Cook et al. 2003)

### ***Características de los cuestionarios identificados***

Las principales características de las 11 escalas evaluadas por EMPRO se presentan en la Tabla 5.1. Los cuestionarios finalmente incluidos se crearon entre los años 1987 y 2003. El número de publicaciones, en los que se basó la evaluación de ellos, varió entre 2 y 30 artículos. Cada escala se compone de un mínimo de 5 hasta un máximo de 30 ítems. Las opciones de respuestas de cuatro cuestionarios (SSI, SST, SDQ-UK, SDQ-NL) cuentan con dos

alternativas en formato Si/No. El cuestionario SPADI es el único que utiliza una escala visual analógica (EVA) en todos sus ítems. Las escalas de ASES-p y SRQ combinan EVA con respuestas tipo Likert. El resto de las escalas cuenta solo con preguntas tipo Likert. Las puntuaciones totales de las escalas estudiadas pueden llegar hasta un máximo de 100 puntos. Aun así, el rango de las puntuaciones totales presenta cierta variación. También varía la interpretación de las puntuaciones mínimas y máximas obtenidas. En algunos casos a mayor puntuación mejor estado del hombro, mientras que en otros puntuaciones más altas significan un peor estado.

Siete de los cuestionarios miden una única dimensión y cuatro dividen sus preguntas en sub-escalas, asignando distintos puntos a cada una de ellas. Concretamente, los cuestionarios que realizan dicha división son: SPADI, ASES-p, PSS y SRQ. El cuestionario SPADI evalúa las dimensiones de *dolor* y *función* con 5 y 8 ítems respectivamente. Cada componente de la escala recibe de 0-100 puntos. La puntuación total es la media de los dos componentes y a mayor puntuación peor función. El cuestionario ASES-p, evalúa las mismas dimensiones que SPADI. En este caso el dolor se basa en un único ítem, que pregunta sobre el nivel de dolor del hombro afectado el día de la encuesta. La función se valora con 10 ítems sobre actividades de la vida diaria. Cada uno de los dos componentes recibe un máximo de 50 puntos y el total de la escala es la suma de las dos sub-escalas. Puntuaciones más altas indican mejor estado de hombro. El cuestionario PSS, aparte del dolor y la función, también evalúa la satisfacción de los pacientes con su propia función del hombro y asigna un máximo de 30, 60 y 10 puntos a cada una de ellas, respectivamente. La puntuación total es la suma de las tres dimensiones. Por último el SDQ contempla un total de cinco sub-escalas: evaluación global del problema, dolor, actividades de la vida, actividades de ocio, satisfacción. El SRQ recoge además el área en que desea mejorar el paciente después de una intervención, aunque dicho apartado no recibe puntuación. Cada sub-escala recibe puntuaciones máximas de 15 a 40 puntos (L'Insalata et al. 1997).

**Tabla 5.1:** Principales características de las 11 escalas evaluadas por EMPRO

Escales	Autor (año)	Artículos EMPRO	Dimensiones (n ítems)	Respuestas y puntuación total	Objetivo	Patología	Min.
SSI	Patte: (1987)	2	Una (30)	Sí/No Puntos: 0-100	Problemas funcionales y AVD	Varias	7
SPADI	Roach (1991)	26	Dolor (5) Función (8)	EVA Puntos: 0-100 (mejor-peor)	Medir dolor y discapacidad	Hombro doloroso	5-8
SST	Lippitt (1993)	12	Una (12)	Sí/No; puntos: 0-12	Medir función del hombro	Varias	<3
SDQ-UK	Croft (1994)	3	Una (22)	Sí/No Puntos: 0-100 (mejor-peor)	Restricciones en las AVD	Hombro doloroso	n.i.
ASES-p	Richards (1994)	30	Dolor (1) Función (10)	EVA & 4-puntos Likert Puntos: 0-100	Valoración de forma estandarizada	Varias	<5
OSS	Dawson (1996)	17	Una (12)	5-puntos Likert Puntos: 12-60 (mejor-peor)	Valorar resultado de operación	Varias menos inestabilidad	<4
SRQ	L'Insalata (1997)	6	Global (1), Dolor (4), AVD (6), Trabajo (5), Tiempo Libre (3), Satisfacción (1), Mejora (1)	EVA & 5-puntos Likert; Puntos: 17-100	Valorar los síntomas y la función del hombro	Varias	5-10
SSRS	Kohn (1997)	3	Una (5)	5-puntos Likert; Puntos: 0-35	Valorar dificultades de función y AVD	Varias	<3
PSS	Leggin (1999)	5	Dolor (3), Función (20), Satisfacción (1)	10-puntos Likert; Puntos: 0-100	Creación de una herramienta específica para el hombro	Varias	<10
SDQ-NL	van der Heijden (2000)	6	Una (16)	Sí/No Puntos: 0-100 (mejor-peor)	Valorar limitaciones en función para pacientes de ensayos clínicos	Trastornos de tejidos blandos del hombro	3
FLEX-SF	Cook (2003)	2	Una (15)	6-puntos Likert; Puntos: 0-60	Crear escala que combina precisión con baja carga al entrevistado	Varias	n.i.

EMPRO: Evaluating the Measurement of Patient Reported Outcomes; EVA: escala visual analógica; AVD: Actividades de la vida diaria; Puntos: rango de puntuación total; salvo si se especifica lo contrario, el rango de las puntuaciones globales indica peor-mejor estado de salud respectivamente. n ítems: número de ítems. Min.: tiempo, en minutos, en el que se rellena la escala; n.i. indica falta de información

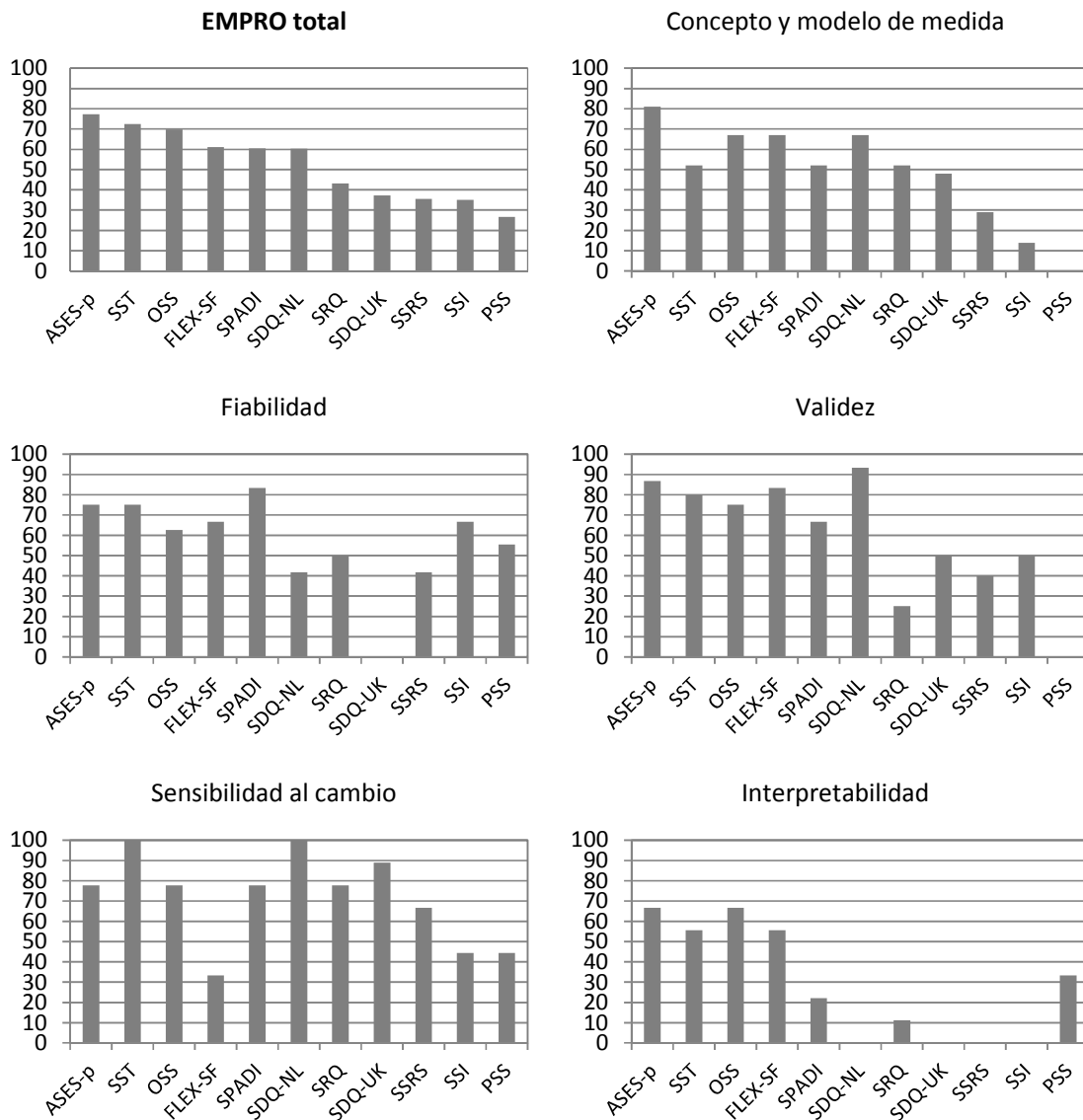
### **Puntuaciones asignadas a los once instrumentos**

Las puntuaciones EMPRO basadas en las evaluaciones consensuadas se presentan en la Tabla 5.2. En ella se incluyen las puntuaciones por ítem, atributo y totales para cada cuestionario. La Figura 5.2 presenta las puntuaciones totales y de los 6 atributos en los que se basa la estimación global. Las puntuaciones totales oscilaron entre 26.7 y 77.4 puntos. Por encima de 50 puntos y por lo tanto resultados aceptables, presentaron los cuestionarios ASES-p, SST, OSS, FLEX-SF, SPADI y SDQ-NL, mientras que los otros cinco cuestionarios obtuvieron valores <45 puntos.

El atributo de *concepto y modelo de medida* obtuvo puntuaciones entre 14.3 y 81 puntos, con los cuestionarios ASES-p (81 puntos), OSS, FLEX-SF y SDQ-NL (66.7 puntos cada uno), alcanzando los valores más altos en este atributo. El PSS no obtuvo puntuación por falta de información. El atributo de la *fiabilidad* presentó menor variación en cuanto a las puntuaciones obtenidas, con un rango entre 41.7 hasta 83 puntos. Este atributo no se pudo evaluar para el SDQ-UK, ya que no se encontró información sobre la *consistencia interna* y la *reproducibilidad* de esta escala. La *validez* no se pudo evaluar para el cuestionario PSS. Por otro lado los cuestionarios SRQ y SSRS recibieron puntuaciones por debajo de los 50 puntos en este atributo. El resto de los instrumentos recibieron entre 50 y 86.7 puntos, con ASES-p y FLEX-SF obteniendo las puntuaciones más elevadas. Ocho de los 11 cuestionarios obtuvieron puntuaciones entre 66.7 y 100 puntos en el atributo de la *sensibilidad al cambio*. La peor puntuación se obtuvo por FLEX-SF (33.3 puntos), mientras que SSI y PSS ambos obtuvieron 44.4 puntos. Finalmente, el atributo de la *interpretabilidad* fue el peor valorado de todos. En relación a este atributo, no se obtuvo información para los instrumentos SDQ-NL, SDQ-UK y SSRS.

Las puntuaciones de los otros 7 cuestionarios oscilaron entre 0 y 66.7 puntos recibidos por SSI y ASES-p respectivamente. Los cuestionarios PSS y SDQ-UK, disponían información solo en 3 de los 5 atributos que componen la puntuación global de EMPRO. Las escalas SDQ-NL y SSRS disponían información en 4 de los 5 atributos.





**Figura 5.2:** Puntuaciones EMPRO totales y por atributo para los 11 escalas evaluadas. Las puntuaciones oscilan de 0 a 100 puntos, representando peor y mejor valor respectivamente.

El cuestionario con la mayor puntuación global fue el ASES-p con 77.4 puntos. Dicho cuestionario obtuvo puntuaciones mayores de 70 puntos en cuatro atributos, siendo la *interpretabilidad* su atributo peor valorado. El SST fue el segundo cuestionario mejor evaluado, con una puntuación total EMPRO >70 puntos. Los atributos de *fiabilidad*, *validez*, y *sensibilidad al cambio* recibieron puntuaciones altas ( $\geq 75$  puntos), mientras que el *concepto y modelo de medida*, y la *interpretabilidad* fueron peor evaluados que

los de ASES-p. La escala con la peor evaluación global fue el PSS con una puntuación de 26.7 puntos, basada en los atributos de *fiabilidad*, *sensibilidad al cambio* e *interpretabilidad*. No hubo información suficiente para poder evaluar el *modelo conceptual* y la *validez* de la misma. Los atributos de *facilidad de uso* y *formas alternativas de administración* no se contemplan en la puntuación global de EMPRO. En cuanto a la *facilidad de uso*, las puntuaciones asignadas oscilaron entre 11.1 (PSS) hasta 100 puntos (SDQ-NL). Finalmente, información sobre *formas alternativas de administración* se encontró solo en dos casos, para las escalas FLEX-SF y SPADI que recibieron 66.7 y 83.3 puntos respectivamente.

**Tabla 5.2:** Puntuaciones EMPRO asignadas a los once cuestionarios de CVRS para la patología del hombro, evaluados en este estudio

Atributo	ASES-p	SST	OSS	FLEX-SF	SPADI	SDQ-NL	SRQ	SDQ-UK	SSRS	SSI	PSS
<b>Concepto y modelo de medida</b>	<b>81</b>	<b>52.4</b>	<b>66.7</b>	<b>66.7</b>	<b>52.4</b>	<b>66.7</b>	<b>52.4</b>	<b>47.6</b>	<b>28.6</b>	<b>14.3</b>	
Concepto de medida descrito	++++	++++	++++	++++	++++	++++	++++	++++	++	++++	++++
Selección de ítems justificada	++++	++	++	++++	++	++++	++	++++	++	-	-
Razonamiento sobre dimensionalidad	++++	++	++	+++	++	++	++	++	++	-	-
Participación población diana	-	++	++++	++++	+	++	++++	+++	-	-	-
Variabilidad descrita y adecuada	++++	++++	++++	-	+++	+++	++	++	++	+	++
Escala de medida descrita y justificada	++++	++	++	+++	++	++	++	+	++	+	-
Derivación de la puntuación descrita	++++	++	+++	++	++++	++++	++	++	++	+	+++
<b>Fiabilidad: puntuación global</b>	<b>75</b>	<b>75</b>	<b>62.5</b>	<b>66.7</b>	<b>83.3</b>	<b>41.7</b>	<b>50</b>		<b>41.7</b>	<b>66.7</b>	<b>55.6</b>
<b>Fiabilidad: consistencia interna</b>	<b>75</b>	<b>58.3</b>	<b>62.5</b>	<b>66.7</b>	<b>83.3</b>	<b>41.7</b>	<b>50</b>				<b>55.6</b>
Métodos de recolección descritos	++++	+++	+++	+++	++++	+++	++	-	-	-	++++
Cronbach/KR-20 adecuados	++++	+++	++++	++++	++++	++++	+++	-	-	-	+++
Se proporcionan estimaciones TRI	-	+++	-	++	+++	-	-	-	-	-	+
Pruebas en diferentes poblaciones	++++	++	NA	NA	+++	-	++++	-	-	-	NA
<b>Fiabilidad: reproducibilidad</b>	<b>75</b>	<b>75</b>	<b>58.3</b>	<b>58.3</b>	<b>66.6</b>		<b>50</b>		<b>41.7</b>	<b>66.7</b>	<b>50</b>
Métodos de recolección descritos	++++	++++	+++	++	+++	-	++	-	++	+++	++
Test-retest e intervalos adecuados	++++	++++	+++	++++	++++	-	++++	++++	+++	++++	+++
Coefficientes de reproducibilidad adecuados	++++	++++	++++	++++	++++	-	+++	-	+++	+++	++++
Se proporcionan estimaciones TRI	-	-	-	-	-	-	-	-	-	-	-
<b>Validez</b>	<b>86.7</b>	<b>80</b>	<b>75</b>	<b>83.3</b>	<b>66.7</b>	<b>93.3</b>	<b>25</b>	<b>50</b>	<b>40</b>	<b>50</b>	
Validez de contenido adecuada	+++	++	+++	++++	++	++++	++	++	++	+	-
Validez de constructo/criterio adecuada	++++	++++	+++	+++	+++	++++	++	+++	++	++	+++
Composición de la muestra descrita	++++	++++	+++	+++	+++	++++	+	+++	++	+++	-
Hipótesis previas descritas	+++	+++	++++	++++	++++	+++	++	++	++	++++	+++
Justificación validez de criterio	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Pruebas en diferentes poblaciones	++++	++++	NA	NA	+++	++++	NA	NA	+++	NA	NA
<b>Sensibilidad al cambio</b>	<b>77.8</b>	<b>100</b>	<b>77.8</b>	<b>33.3</b>	<b>77.8</b>	<b>100</b>	<b>77.8</b>	<b>88.9</b>	<b>66.7</b>	<b>44.4</b>	<b>44.4</b>
Métodos adecuados	++++	++++	+++	++	+++	++++	++	+++	++	+++	++
Magnitud de cambio descrita	++++	++++	++++	++	++++	++++	++++	++++	+++	+++	++++
Comparación de grupos estables e inestables	++	++++	+++	++	+++	++++	++++	++++	++++	+	+
<b>Interpretabilidad</b>	<b>66.7</b>	<b>55.6</b>	<b>66.7</b>	<b>55.6</b>	<b>22.2</b>		<b>11.1</b>			<b>0</b>	<b>33.3</b>
Justificación de criterios externos	+++	+++	+++	+++	++	-	++	-	-	+	++
Descripción del modo de interpretación	+++	++	++	++	++	-	+	-	-	+	++
Cómo presentar resultados descrito	+++	+++	++++	+++	+	-	+	-	-	-	++
<b>Puntuación total EMPRO</b>	<b>77.4</b>	<b>72.6</b>	<b>69.7</b>	<b>61.1</b>	<b>60.5</b>	<b>60.3</b>	<b>43.3</b>	<b>37.3</b>	<b>35.4</b>	<b>35.1</b>	<b>26.7</b>

**Tabla 5.2:** continúa

Atributo	ASES-p	SST	OSS	FLEX-SF	SPADI	SDQ-NL	SRQ	SDQ-UK	SSRS	SSI	PSS
<b>Facilidad de uso</b>											
<b>Facilidad de uso I: para el entrevistado</b>	<b>55.6</b>	<b>88.9</b>	<b>88.9</b>		<b>22.2</b>	<b>100</b>	<b>22.2</b>	<b>77.8</b>	<b>66.7</b>	<b>11.1</b>	<b>11.1</b>
Habilidad y tiempo necesario	+++	++++	+++	-	++	++++	++	++++	++++	++	++
Impacto en el entrevistado	++	++++	++++	+++	++	++++	++	+++	++++	+	+
Circunstancias no apropiadas	+++	+++	++++	-	-	++++	+	+++	-	-	-
<b>Facilidad de uso II: carga de administración</b>	<b>91.7</b>	<b>41.7</b>	<b>66.7</b>	<b>16.7</b>	<b>50</b>	<b>100</b>	<b>33.3</b>	<b>58.3</b>	<b>50</b>	<b>25</b>	<b>75</b>
Recursos necesarios	+++	+++	++++	++	++++	++++	+	+++	++++	+	++++
Tiempo necesario	++++	-	-	-	-	++++	++	++++	-	++++	++++
Formación y experiencia necesarias	++++	-	++++	-	-	++++	+	+++	-	+	-
Carga asociada al cálculo de las puntuaciones	++++	++++	+++	++	++++	++++	++++	-	++++	+	++++
<b>Formas alternativas de administración</b>											
Características métricas y de uso	NA	NA	NA	<b>66.7</b>	<b>83.3</b>	NA	NA	NA	NA	NA	NA
Comparación de las formas alternativas	NA	NA	NA	+++	+++	NA	NA	NA	NA	NA	NA

Una puntuación de ++++ indica 4 (muy de acuerdo); +++ es 3; ++ es 2; + es 1; - es No información y NA, No aplicable; TRI: Teoría de la respuesta al ítem. Las puntuaciones EMRPO oscilan de 0 a 100 puntos, representando el peor y mejor valor respectivamente.

## 5.4 Discusión

En el presente estudio se evaluaron las cualidades psicométricas de varias escalas auto-administradas de CVRS, especialmente diseñadas para la evaluación de la patología del hombro. Según las evaluaciones estandarizadas de 22 expertos en medidas de este tipo, los cuestionarios mejor puntuados por EMPRO fueron los ASES-p, SST y OSS, mientras que resultados aceptables se encontraron para las escalas de FLEX-SF, SPADI y SDQ-NL. Todas las escalas evaluadas en el presente trabajo, son relativamente cortas y fáciles de administrar. Asimismo y considerando siempre sus características particulares, cada una de ellas se podría implementar en proyectos de investigación, según los objetivos planificados.

El ASES-p (Richards et al. 1994) obtuvo el mejor resultado global y recibió >70 puntos en casi todos los atributos implementados en el cálculo total de EMPRO, salvo en el atributo de la *interpretabilidad*. Ninguno de los ítems de este atributo consiguió la máxima puntuación. Las diferencia clínicamente importantes se han explorado en una muestra anterior de tamaño reducido (Michener et al. 2002). El SST (Lippitt et al. 1993) obtuvo puntuaciones altas en los atributos de *fiabilidad*, *sensibilidad al cambio*, y *validez*. Por otro lado, la puntuación conseguida en el *concepto y modelo de medida* fue baja, ya que la información relacionada con el proceso de su desarrollo, participación de la población diana y el nivel de medida de la escala, no fue suficiente. La *interpretabilidad* de esta escala también se situó muy cerca del punto de corte de la puntuación aceptable, ya que no se encontró suficiente información sobre la interpretación de las puntuaciones frente a diferentes "*anclajes*" (Godfrey et al. 2007). El cuestionario OSS estaba entre los 3 mejores en cuanto a los atributos de *concepto y modelo de medida*, e *interpretabilidad*. Sus puntuaciones de *validez* y *sensibilidad al cambio* también fueron aceptables, mientras que su *fiabilidad* recibió menos de 70 puntos, debido a que ciertos aspectos relacionados con la recogida de datos y tiempo transcurrido entre test-retest no fueron presentados adecuadamente. Dado que los instrumentos ASES-p, SST y OSS son similares en contenido, número de ítems y tiempo de administración, la selección del más apropiado se podría basar en diferentes

consideraciones, como por ejemplo la dimensionalidad de los mismos, el número de ítems, sus alternativas de respuesta etc., según el contexto que deseamos aplicarlos. El ASES-p está compuesto por las dimensiones de dolor y función, que se miden por una escala visual analógica (EVA) y por respuestas tipo Likert. Los cuestionarios SST y OSS son unidimensionales y su puntuación total se deriva por respuestas dicotómicas y tipo Likert respectivamente.

Los cuestionarios FLEX-SF, SPADI y SDQ-NL obtuvieron alrededor de 60 puntos totales. Sus resultados fueron aceptables en la mayoría de los atributos evaluados, pero cada una de las escalas recibió <50 puntos en uno de ellos. Concretamente, el FLEX-SF (Cook et al. 2003) obtuvo una puntuación baja en *sensibilidad al cambio*, ya que a pesar de la presentación de los efectos estandarizados, no estaba claramente explicado el método de cálculo de los mismos. El SPADI (Roach et al. 1991) obtuvo un valor bajo en el atributo de la *interpretabilidad*. Aunque se trata de una escala ampliamente utilizada, los datos sugieren que este atributo requiere más exploración. Finalmente más investigación hace falta acerca del atributo de la *fiabilidad* de SDQ-NL (van der Heijden et al. 2000). Aun así, cabe mencionar que debido a su excelente *sensibilidad al cambio* y *carga para el entrevistado* (respuestas binarias y tiempo de cumplimentación 3 minutos), dicha escala podría ser una buena opción para medir *sensibilidad al cambio* en proyectos longitudinales.

Nuestro proyecto tiene ciertas limitaciones. Primero, las evaluaciones EMPRO aquí presentadas se basan en la información encontrada en la revisión sistemática de la literatura, realizada solo en la librería electrónica de PubMed. Aunque PubMed es una de las librerías más importantes en el ámbito de la salud, es posible que no se hayan identificado todos los artículos disponibles sobre el tema. Esperamos que la sensibilidad de la búsqueda, combinando palabras clave y términos MeSH, así como la posterior búsqueda manual de las referencias bibliográficas, de todos los artículos incluidos, haya minimizado este problema. Una segunda limitación es el hecho que las evaluaciones EMPRO se ven afectadas por la cantidad y calidad de la información encontrada. Falta de información, penaliza los resultados globales de EMPRO. En este estudio, la mayoría de las escalas fueron penalizadas por falta o información inadecuada sobre

*interpretabilidad*. Este dato subraya la necesidad de trabajar más en este atributo y proporcionar información más adecuada acerca de la interpretación de las puntuaciones de cada escala. Finalmente las evaluaciones EMPRO podrían haber sido afectadas por el nivel de conocimiento y experiencia de los evaluadores. Las instrucciones detalladas de la propia herramienta, la evaluación por pares y el consenso de todas las evaluaciones deberían haber atenuado también este sesgo. En esta evaluación, las adaptaciones culturales en otros idiomas no se han valorado por separado, sino que se han considerado parte de la evidencia global de cada instrumento. Consideramos que la evaluación conjunta de estos datos refleja mejor y de manera más amplia las cualidades psicométricas de las escalas incluidas.

Hasta donde conocemos éste es el primer estudio, en realizar una revisión sistemática sobre las escalas de CVRS para la patología del hombro y evaluar de modo estandarizado la evidencia disponible mediante la herramienta EMPRO. Nuestros resultados pueden ser de interés tanto en el ámbito de la práctica clínica, como para el ámbito de la investigación en escalas auto-administradas. Para las necesidades de este proyecto nos centramos solo en escalas auto-cumplimentadas por los propios pacientes. No se consideraron escalas que requieren la valoración de un especialista, como la escala de Constant; tampoco escalas diseñadas para patologías concretas o grupos específicos de pacientes.

La evidencia disponible sugiere que los cuestionarios ASES-p, SST y OSS estarían entre las mejores opciones para medir la funcionalidad en pacientes con patologías de hombro. Estas escalas son fiables, válidas y sensibles al cambio. Asimismo, su *concepto y modelo de medida*, la *interpretabilidad* y *facilidad de uso* son aceptables. La administración de las escalas FLEX-SF, SPADI y SDQ-NL también se podría considerar ya que dichos instrumentos presentan resultados aceptables en la mayoría de los atributos evaluados.





## **6. Sección auto-cumplimentada del cuestionario ASES: validación al castellano aplicando métodos de la teoría clásica y el modelo Rasch**

Los datos presentados en este capítulo se han publicado en la revista "Health and Quality of Life Outcomes": Patient self-report section of the ASES questionnaire: a Spanish validation study using classical test theory and the Rasch model. Health Qual Life Outcomes. 2016 Oct 18; 14(1):147. Vrotsou K, Cuéllar R, Silió F, Rodríguez MÁ, Garay D, Busto G, Trancho Z, Escobar A.

## 6.1 Introducción

Las patologías del hombro se encuentran entre los más habituales problemas de salud (Vicente-Herrero et al. 2009, American Academy of Orthopaedic Surgeons 2013). Estas dolencias limitan las actividades de la vida diaria (Largacha et al. 2006), aumentan el absentismo laboral (Kuijpers T et al. 2006) y afectan la calidad de vida de los pacientes (Paananen et al. 2011). En la literatura científica se encuentran muchas escalas auto-administradas que evalúan la CVRS para pacientes con patologías de hombro. Tras una revisión sistemática y evaluación estandarizada de la evidencia disponible, realizada por nuestro grupo y descrita en el capítulo anterior (Schmidt et al. 2014), se decidió que la parte auto-cumplimentada de la escala *American Shoulder and Elbow Surgeons score* (ASES-p) era la más adecuada para adaptar y validar al castellano.

El cuestionario ASES se publicó por primera vez en 1994 y sus propiedades psicométricas han sido estudiadas en pacientes con diferentes patologías del hombro (Kocher et al. 2005, Michener et al. 2002). Evalúa dos componentes el nivel del dolor y la dificultad en realizar ciertas actividades de la vida diaria (AVD). El ASES-p es una escala ampliamente utilizada que ha sido culturalmente adaptada y validada en varios idiomas (Yahia et al. 2011, Goldhahn et al. 2008, Piitulainen et al. 2014, Padua et al. 2010, Knaut et al. 2010, Celik et al. 2013). El capítulo actual presenta los datos de la validación de esta escala al castellano. Dicha validación se realizó aplicando, entre otros, los métodos de análisis factorial confirmatorio (AFC) y el modelo Rasch.

## 6.2 Metodología

### ***Criterios de Inclusión y exclusión***

Los participantes de este proyecto se reclutaron en los servicios de las consultas de traumatología de cinco hospitales públicos del País Vasco. Los centros participantes fueron los hospitales universitarios de Donostia, Basurto y Cruces, el hospital de Galdakao-Usansolo y el hospital de Mendaró. Se reclutaron pacientes  $\geq 18$  años de edad

que acudían a la consulta y entraban en la lista de espera para recibir una intervención quirúrgica o conservadora en el hombro afectado. Se reclutaron solo sujetos con alto nivel de castellano escrito y hablado. Se excluyeron pacientes con intervenciones anteriores en el hombro afectado y pacientes con deterioro cognitivo que podría haber comprometido su capacidad en contestar a las preguntas de los cuestionarios administrados. Todos los sujetos reclutados firmaron el consentimiento informado.

### ***Recogida de datos***

#### ***Datos recogidos por los especialistas***

Los traumatólogos realizaron, en sus consultas habituales, la prueba funcional de Constant-Murley (CMS). Los componentes de dolor y AVD, de esta escala, se respondieron por los propios pacientes, mientras que los especialistas evaluaron el rango de movimiento (RM) y la fuerza de la articulación afectada. Para esta última se utilizaron diferentes tipos de pesas, según los recursos de cada servicio. Para las necesidades de este proyecto, las puntuaciones de CMS se derivaron de tres maneras distintas: considerando todos los ítems, según las indicaciones de la publicación original (CMS<sub>O</sub>); ajustando por edad y sexo, según las recomendaciones más recientes del propio autor original (CMS<sub>R</sub>) y finalmente excluyendo el apartado de la fuerza (CMS<sub>NS</sub>).

#### ***Datos auto-cumplimentados por los participantes***

El cuestionario ASES-p, las escalas SF-36 y Barthel, así como preguntas socio-demográficas, el estado civil, información sobre consumo de fármacos, las patologías adicionales, las expectativas de los pacientes por el tratamiento a recibir y otros datos de interés se rellenaron por los reclutados en su propio domicilio. Los datos relacionados con las intervenciones y tratamientos administrados se recogieron por la información incluida en las historias clínicas de los participantes.

### ***Seguimiento***

Tanto la prueba funcional como la batería de preguntas se cumplimentaron dos veces, en el momento basal y al final del seguimiento. Para los sujetos intervenidos quirúrgicamente, la segunda valoración se realizó a los 6 meses de la intervención y para los sujetos con intervención conservadora, a los 3 meses de la misma. Los periodos de tiempo correspondientes se establecieron siguiendo las indicaciones de los traumatólogos colaboradores.

### ***Coordinación de la recogida***

Una técnico de investigación, con experiencia en este tipo de proyectos y contratada para realizar la gestión de los datos del presente estudio, coordinó todas las tareas relacionadas con la recogida de datos y el estricto seguimiento de los participantes. En caso de no respuesta al primer envío se les enviaba una carta recordatoria y a continuación se localizaban por teléfono los que seguían sin responder. Si después de estos intentos el sujeto aun no había contestado a la batería de preguntas enviadas al domicilio, se consideraba dato perdido. Los datos relacionados con la intervención administrada y otros datos médicos se recogieron de la historia clínica. Solo un hombro por sujeto se considero en este estudio.

### ***Tamaño Muestral***

La teoría psicométrica sugiere reclutar alrededor de 10 pacientes por cada ítem del cuestionario a validar (Brown 2006). Considerando que el ASES-p se compone de 11 y esperando que un 30% de respuestas perdidas, para cumplir los requisitos del tamaño muestral haría falta reclutar un mínimo de N=160 sujetos.

### **Análisis estadístico**

Los datos categóricos se presentan con frecuencias y porcentajes (%) y los continuos con medias y desviaciones estándar (DE). Asociaciones entre variables categóricas se estudian con el chi-cuadrado y diferencias entre medias se realizan con el t-test de Student para medias emparejadas o independientes según el contexto. Ciertos análisis, relacionados con la consistencia interna y la estructura de ASES-p se realizaron dos veces, una incluyendo todos los 11 ítems de la escala y otra excluyendo el ítem de dolor. El ítem del dolor se utilizó por un lado en formato reverso es decir, el “dolor máximo” y “no dolor” se asignaron 0 y 10 puntos respectivamente (en vez de los 10 y 0 que la escala asigna). La transformación se realizó para que este ítem fuese en el mismo sentido que los otros ítems de la escala, donde las respuestas “0=imposible” y “3=sin dificultad”. Esta transformación se utilizó en la derivación de las correlaciones y en el modelo de AFC. El mismo ítem se transformó, además, en una variable de 4 categorías. En este caso, valores 0-1 se consideraron como respuesta “3=sin dolor”; valores 2-5 como “2=con cierto dolor”; valores 6-8 como “1=con mucho dolor” y por último valores de 9-10 como “0=máximo dolor”. Los puntos de corte considerados en esta categorización se decidieron tras estudiar las respuestas de este ítem, frente a las respuestas en los ítems de dolor y salud general de SF-36. Esta transformación se usó para la estimación de alpha de Cronbach y en los modelos Rasch.

### **Consistencia Interna**

El alpha de *Cronbach*, así como con las correlaciones entre ítems e ítem-puntuación total, sin considerar el respectivo ítem, se implementaron para evaluar la consistencia interna de la escala. Se consideraron aceptables valores  $\geq 0.70$  y  $\geq 0.30$  respectivamente (Terwee et al. 2007). El índice de Cronbach, se calculó dos veces: considerando solamente los 10 ítems de la función y todos los 11 ítems de la escala.

### **Validez estructural**

La *validez estructural* se evaluó mediante modelos de AFC y el modelo Rasch. El hecho de que la dimensión del dolor de ASES-p contiene un único ítem no permitió realizar distintos modelos para una de las dos dimensiones de esta escala. Por este motivo, se consideraron dos modelos. El primero con todos los 11 ítems de la escala y el segundo con los 10 ítems de la sub-escala de la función. Ambos modelos estudian la posibilidad de un único factor latente. Se realizaron modelos de AFC para datos categóricos implementando el método de Unweighted Least Squares (ULSWV). Este método se recomienda para datos ordinales y muestras <200 sujetos. Se consideraron aceptables cargas factoriales  $\geq 0.40$  (Brown 2006). Asimismo se estudiaron los siguientes índices de ajuste: Root Mean Square Error of Approximation (RMSEA), con valores aceptables <0.08 y los Tucker-Lewis Index (TLI) y comparative fit index (CFI) con valores >0.90. Los valores residuales e índices de modificación se examinaron en cada caso. La unidimensionalidad de la escala se estudió además mediante el modelo Rasch, considerando el dolor como ítem categórico. Se implementó el *Andrich Rating Scale Model* para ítems con respuestas politómicas. Se calculó la *dificultad* (en logit) de todos los ítems, se comprobó la bondad de ajuste de la escala con los índices de Infit y Outfit. Los valores de estos últimos deberían situarse entre 0.6 y 1.4. Ítems con valores <0.5 no aportan mucha información a la escala, aunque no distorsionan las medidas proporcionadas. Esta distorsión sí que sucede cuando los respectivos valores son >2.0, por lo que estos ítems se consideran no productivos y se aconseja su eliminación de la escala de interés. Asimismo, se estudió la fiabilidad de sujetos e ítems (*person and ítem reliability*) y los índices de separación (*separation statistics*). Para estos análisis los valores deseables serían > 0.8 y > 2.0 respectivamente (Bond & Fox 2007). De manera adicional, la unidimensionalidad de la escala se comprobó analizando los residuos del modelo Rasch vía análisis de componentes principales. Valores eigenvalue <3 apoyan la unidimensionalidad de la escala.

### **Validez convergente & discriminante**

La *validez convergente* se estudió mediante las correlaciones de la escala ASES-p con la escala CMS, la parte física de la escala SF-36 (PCS) y el cuestionario de Barthel. La *validez discriminante* se estudió mediante la correlación con las dimensiones y el componente mental de SF-36 (MCS). Nuestra hipótesis previa fue que la correlación de ASES-p con estas variables serían más bajas que las correlaciones observadas con las variables físicas de SF-36 y las otras escalas utilizadas. Se implementaron las correlaciones de Spearman ( $r_s$ ).

### **Validez de grupos conocidos**

La validez de grupos conocidos de ASES-p se estudió frente a las escalas CMS<sub>0</sub>, PCS y Barthel, tras transformarlas en variables categóricas. Las comparaciones en variables con 3 grupos ordinales se realizaron con el estadístico no paramétrico de Jonckheere-Terpstra, que comprueba si existe una tendencia de aumento (o disminución) de los valores del cuestionario, entre los consecutivos grupos. Comparaciones entre dos grupos se realizaron con el t-test de Student para muestras independientes.

### **Sensibilidad al cambio**

En el segundo momento del estudio (post tratamiento) se les preguntó a los sujetos, si según su opinión, su rango de movimiento y su capacidad en realizar sus AVD (de manera global, sin indicar ninguna actividad concreta), habían mejorado en comparación con el momento basal. Aquellos sujetos que dieron una respuesta positiva en ambas preguntas se consideraron como “mejorados”. Nuestra hipótesis fue, que estos pacientes presentarían diferencias pre-post, en las puntuaciones de ASES-p, más elevadas que el resto de los sujetos. Se calcularon tres tamaños de efecto. El efecto estandarizado (SES), la respuesta media estandarizada (SRM) y la respuesta media estandarizada, ajustada para datos emparejados (SRM<sub>Adj</sub>) (Husted et al. 2000).

Según la definición de Cohen, valores  $>0.80$  indicarían efectos grandes, mientras que valores de 0.50 y 0.20 se considerarían efectos medianos y pequeños respectivamente (Cohen 1992). Los diferentes efectos se estudiaron también considerando tres grupos de intervención: quirúrgica, infiltración y otra.

### **6.3 Resultados**

#### ***Datos basales***

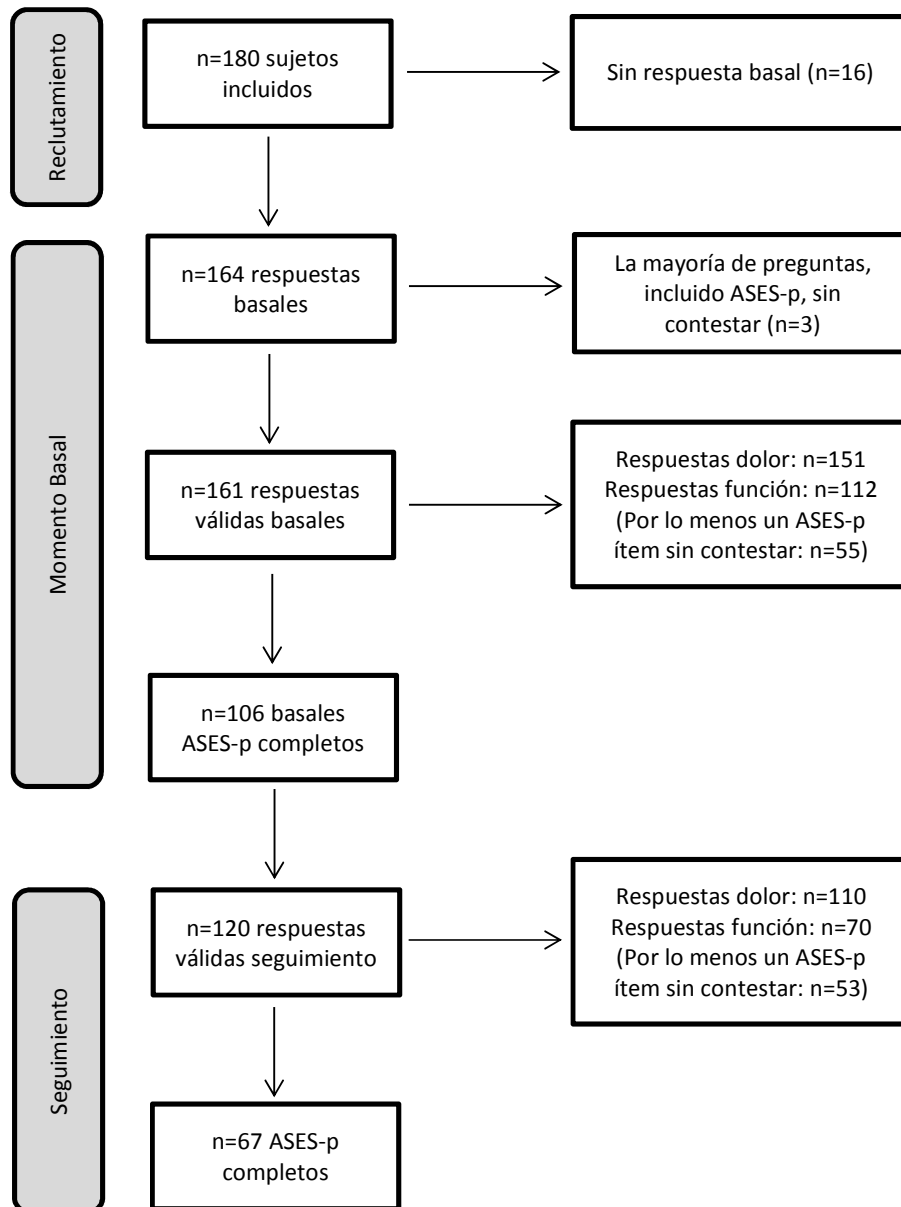
El reclutamiento duró entre Mayo 2012 hasta Noviembre 2013 y el seguimiento finalizó en Junio de 2014. Se reclutaron  $n=180$  sujetos que cumplían los criterios de inclusión. De ellos,  $n=164$  respondieron a la batería de preguntas basales, pero  $n=3$  no contestaron a ninguna las preguntas de ASES-p. Por lo tanto, en el momento basal se recibieron  $n=161$  contestaciones válidas.

La mayoría de los ítems de la escala ASES-p, presentaba entre 1 y 6% de datos perdidos, mientras que el último ítem de la escala “realizar su deporte habitual” no se contestó por el 23% de los participantes. En total,  $n=151$  sujetos respondieron al ítem del dolor y  $n=112$  a todos los ítems de la sub-escala de la función. En el momento basal se obtuvieron  $n=106$  respuestas completas de ASES-p, mientras que el número de respuestas completas a los 6 meses fue  $n=67$  (Figura 6.1).

La edad comprendida de los participantes fue entre 22 y 82 años. La mitad de ellos eran mujeres (51%). La mayoría padecía síndrome subacromial con rotura del manguito rotador (64%), mientras que el lado afectado fue sobre todo el derecho (68%). El 35% de los individuos tenía un trabajo manual. El 33% había cogido en los últimos 5 años por lo menos una baja laboral, debido a su problema de hombro. Varios sufrían además por problemas de espalda (52%), cervicales (47%) o problemas de las extremidades inferiores (Tabla 6.1). En todos los ítems de la escala, la opción “imposible” fue seleccionada por 5-56% de los participantes, mientras que la opción, “sin dificultad” se seleccionó por 5-31% de ellos. La media puntuación de ASES-p, en el hombro afectado,



fue 46.5 (DE: 23.0) puntos. No se observaron efectos suelo y techo para la puntuación total de la escala ya que solo 1% presentó el mínimo y máximo valor, siendo estos 5 y 100 puntos, respectivamente.



**Figura 6.1:** Diagrama de flujo con datos de participación en el proyecto de validación de ASES-p.

**Tabla 6.1:** Características de la muestra y valores basales de ASES-p

<b>Variables</b>	<b>N=161</b>
<b>Edad en años; media (DE)</b>	59.7 (11.8)
<b>Sexo</b>	
Mujer	82 (51)
Hombre	79 (49)
<b>Hombro afectado</b>	
Derecho	110 (68)
Izquierdo	51 (32)
<b>IMC</b>	
<25	50 (32)
25-29.9	69 (44)
≥30	39 (25)
<b>Años con patología; mediana (Q1, Q3)</b>	3 (1, 7)
<b>Fumador</b>	
Sí actualmente	31 (19)
Ex fumador/a	48 (30)
Nunca ha fumado	82 (51)
<b>Estado Civil</b>	
Vive con pareja	120 (75)
Vive solo/a	26 (16)
Viudo/a	14 (9)
<b>Nivel de educación</b>	
Primaria o menos	67 (42)
Secundaria	70 (44)
Universitaria o más	22 (14)
<b>Patologías adicionales</b>	
Espalda	84 (52)
Cuello	75 (47)
Extremidad inferior	63 (39)
Extremidad superior	10 (6)
<b>Tipo de trabajo</b>	
Manual	56 (35)
Oficina	26 (16)
Amo/a de casa	31 (20)
Pensionista	36 (23)
Estudiando / en paro	3 (2)
<b>Baja laboral los últimos 5 años</b>	
Sí	53 (33)
No	108 (67)
<b>Diagnostico</b>	
Síndrome subacromial con ruptura MC	103 (64)
Síndrome subacromial sin ruptura MC	13 (8)
Tendinopatía: tendinitis/tendinosis	27 (17)
Inestabilidad: traumática/atraumática	7 (4)
Artrosis	5 (5)
Necrosis avascular	1 (1)
Hombro congelado	2 (1)
<b>ASES-p; media (DE)‡</b>	
Puntuación total (0-100 puntos)	46.5 (22.9)
Dolor (0-50 puntos)	23.8 (12.9)
Función (0-50 puntos)	22.9 (11.9)

Los números son frecuencias (porcentajes), salvo si se indica lo contrario. DE: desviación estándar; Q1, Q3: primero y tercer cuartil. MR: Manguito rotador. Los pacientes pueden tener una o más patologías adicionales. ‡ Información se basa en los datos disponibles: puntuación total n=106; dolor n=151 and función n=112 subjects.

Ochenta y dos sujetos (51%) proporcionaron también respuestas completas para el hombro contralateral. Según ellas, la media puntuación ASES-p para este hombro fue 66.3 (DE: 21.1) puntos y la diferencia entre los dos hombros fue estadísticamente significativa ( $p < 0.0001$ ). De los participantes  $n=66$  se sometieron a una intervención quirúrgica;  $n=28$  a infiltraciones;  $n=8$  sujetos rechazaron la intervención;  $n=2$  estaban todavía pendientes de intervención cuando finalizó el seguimiento y el resto realizó un programa de rehabilitación. Los sujetos que no respondieron al ítem 10 “realizar deporte habitual” tenían una edad media de 64.9 (SE: 10.3) años.

No se encontraron diferencias estadísticamente significativas en términos de edad ( $p=0.348$ ), sexo ( $p=0.339$ ) y las tres versiones de CMS ( $p > 0.900$ ) entre los respondedores y no respondedores de este ítem.

### **Consistencia Interna**

El alpha de Cronbach obtuvo el valor 0.91, tanto consideraron todos los 11 ítems de la escala, como después de excluir el ítem del dolor. Las correlaciones entre ítems fueron  $r_s > 0.30$  y solo el ítem 10. “realizar su deporte habitual” presentó correlaciones más bajas en dos ocasiones; con el ítem del dolor ( $r_s=0.261$ ) y con el ítem 3. “lavarse la espalda” ( $r_s=0.291$ ). Las correlaciones ítem-total de la escala, controlando el solapamiento, se estimaron entre 0.73 y 0.40, siendo el ítem 10 el que presentó el valor más pequeño (Tabla 6.2). Las sub-escalas del dolor y de la función presentaron una correlación sustancial entre ellas  $r_s=0.605$ . Finalmente, el 17% de la muestra obtuvo puntuaciones más altas en la sub-escala de la función ( $>25$  puntos) y relativamente bajas en la sub-escala del dolor ( $<25$  puntos) o viceversa.

### **Validez**

En los AFC todos los ítems de la escala presentaron cargas factoriales  $>0.50$  y los coeficientes TLI y CFI fueron  $>0.90$ . El modelo que consideró todos los ítems obtuvo un RMSEA=0.058 (90%IC: 0.00-0.092), mientras que para el modelo con los 10 ítems de la función se obtuvo un RMSEA=0.085 (90%IC: 0.050-0.118). En este caso y basados en los

resultados de los índices de modificación, se introdujo en el modelo una correlación entre los ítems 9 y 10 (Tabla 6.2).

**Tabla 6.2:** Resultados de AFC y correlaciones ítem-total de la escala

Ítem no.	Descripción	Todos los ítems (n=106)	Ítems de función¥ (n=112)	Correlaciones ítem-total
Dolor ‡ ítem	Intensidad de dolor hoy	0.62	-	0.60
Función				
ítem 1	Ponerse un abrigo	0.83	0.79	0.66
ítem 2	Dormir sobre lado afectado	0.73	0.75	0.58
ítem 3	Lavarse la espalda	0.79	0.79	0.60
ítem 4	Limpiarse tras micción/defecación	0.69	0.74	0.52
ítem 5	Peinarse	0.82	0.82	0.65
ítem 6	Llegar a una balda alta	0.89	0.87	0.73
ítem 7	Levantar 4.5kg por encima del hombro	0.84	0.85	0.62
ítem 8	Lanzar pelota por encima de la cabeza	0.88	0.86	0.70
ítem 9	Realizar trabajo habitual	0.83	0.81	0.66
ítem 10	Realizar deporte habitual	0.53	0.57	0.40
<i>Diagnósticos</i>				
$\chi^2$ ; gl		59.476; 44	61.507; 34	
RMSEA (IC 90%)		0.058 (0.0-0.092)	0.085 (0.050-0.118)	
p-valor RMSEA		0.349	0.052	
TLI; CFI		0.942; 0.954	0.981; 0.975	

AFC: Análisis factorial confirmatorio. ‡ El ítem del dolor se implementó en su formato reverso, con valores 0=intensidad máxima y 10=no dolor. ¥ En este modelo, se permitió una correlación entre los ítems 9 & 10 (carga factorial=0.44). Los resultados del AFC y las correlaciones se basaron en todos los datos disponibles. Las correlaciones ítem-total de la escala se estimaron con el coeficiente de Spearman y han sido controladas por solapamiento. gl: grados de libertad. RMSEA: root mean square error of approximation. IC: intervalo de confianza. TLI: Tucker-Lewis index. CFI: comparative fit Index.

Los respectivos modelos Rasch se presentan en la Tabla 6.3. Con la excepción del ítem 6. “llegar a una balda alta” (en ambos modelos) y el ítem del dolor (en el primer modelo), el resto de los ítems presentó una distancia logit > 0.10, indicando suficiente dispersión en cuanto a la dificultad de las preguntas ASES-p. La exclusión del dolor, en el segundo, modelo no cambio el orden de la dificultad de los ítems, aunque las estimaciones de los ítems 6 y 2. “dormir sobre el lado afectado” fueron iguales. En la mayoría de los ítems, los estadísticos de infit y outfit presentaron valores aceptables.

El ítem 10 tuvo infit y outfit >1.4 en ambos modelos, mientras que el outfit del dolor en el primer modelo fue 1.74. En ningún caso se observaron valores >2, mientras que los estadísticos de separación y fiabilidad también fueron aceptables para ambos modelos.

Finalmente, el análisis de componentes principales de los residuos de los modelos Rasch no indicó la presencia de más factores latentes ya que los eigenvalue fueron de 2.12 y 2.10, respectivamente.

**Tabla 6.3:** Nivel de dificultad, errores estándar, estadísticos de ajuste y orden según dificultad de los ítems en los dos modelos Rasch (n=161)

ítem	Descripción	Todos los ítems ¥					Ítems de función ‡					
		δ (logit)	EE	Infit MNSQ	Outfit MNSQ	Rank	δ (logit)	SE	Infit MNSQ	Outfit MNSQ	Rank	
Función												
ítem 7	Levantar 4.5kg por encima del hombro	1.78	0.13	1.09	0.86	1	1.82	0.14	1.09	0.89	1	
ítem 8	Lanzar pelota por encima de la cabeza	1.07	0.13	0.82	0.72	2	1.09	0.13	0.81	0.73	2	
ítem 3	Lavar la espalda	0.89	0.12	0.90	0.83	3	0.90	0.13	0.90	0.85	3	
ítem 2	Dormir sobre lado afectado	0.37	0.12	0.98	1.01	4	0.35	0.12	1.02	1.10	4	
ítem 6	Llegar a una balda alta	0.36	0.12	0.85	0.79	5	0.35	0.12	0.86	0.82	4	
ítem 10	Realizar deporte habitual	-0.12	0.14	1.81	1.78	6	-0.16	0.14	1.86	1.90	5	
ítem 9	Realizar trabajo habitual	-0.28	0.12	0.73	0.74	7	-0.32	0.13	0.80	0.88	6	
Dolor ‡												
Ítem	Intensidad de dolor hoy	-0.33	0.12	1.26	1.74	8	-	-	-	-	-	
Función												
ítem 1	Ponerse un abrigo	-0.97	0.13	0.71	0.75	9	-1.05	0.13	0.78	0.89	7	
ítem 5	Peinarse	-1.28	0.13	0.79	0.78	10	-1.38	0.13	0.81	0.83	8	
ítem 4	Limpiarse tras micción/defecación	-1.50	0.13	1.19	1.11	11	-1.61	0.14	1.23	1.16	9	
Index												
fiabilidad: persona / ítem			0.90 / 0.98					0.88 / 0.98				
separación: persona / ítem			2.93 / 7.53					2.77 / 8.02				

δ (logit) : nivel de la dificultad de los ítems, según las estimaciones del modelo Rasch; EE: Error estándar; MNSQ: mean square fit index; Rank: orden de dificultad según δ (logit). ‡ En este modelo el ítem de dolor se utilizó como variable con 4 categorías de respuesta: 0-1=No dolor, 2-5=algo de dolor; 6-8= mucho dolor; 9-10=máximo dolor. En este modelo los ítems 2 y 6 obtuvieron la misma estimación δ (logit) y por eso se asignan el mismo orden.

Los datos de validez convergente, discriminantes y de grupos conocidos se presentan en la Tabla 6.4. La puntuación total de ASES-p presentó correlaciones  $r_s > 0.50$ , con todas las versiones de CMS y con los componentes del dolor y AVD de CMS<sub>0</sub>. Las correlaciones de ASES-p fueron más elevadas con las dimensiones físicas y el componente PCS de SF-36, mientras que correlaciones más bajas se observaron con el componente mental MCS y la escala de Barthel.

**Tabla 6.4:** Validez convergente, discriminante y de grupos conocidos de ASES-p con CMS SF-36 y Barthel

Correlación de Spearman		ASES-p total (n=106)	Función (n=112)	Dolor (n=151)
CMS versión				
CMS <sub>O</sub>		<b>0.62</b>	<b>0.59</b>	0.45
CMS <sub>R</sub>		<b>0.54</b>	0.49	0.41
CMS <sub>NS</sub>		<b>0.64</b>	<b>0.59</b>	0.47
CMS <sub>O</sub> componente				
Dolor		<b>0.62</b>	0.44	<b>0.56</b>
AVD		<b>0.57</b>	<b>0.53</b>	0.45
RM		0.46	0.45	0.27
Fuerza		0.37	0.39	0.23
SF-36				
PCS		<b>0.65</b>	<b>0.60</b>	<b>0.52</b>
MCS		0.32	0.29	0.28
Función física		<b>0.59</b>	<b>0.54</b>	0.46
Rol físico		<b>0.60</b>	<b>0.60</b>	0.47
Dolor		<b>0.74</b>	<b>0.67</b>	<b>0.66</b>
Salud general		0.39	0.37	0.27
Vitalidad		<b>0.54</b>	0.44	0.40
Función social		0.45	0.43	0.35
Rol emocional		0.33	0.29	0.34
Salud mental		0.39	0.35	0.35
Barthel		0.31	0.33	0.22
<b>Grupos conocidos</b>				
	<b>n</b>			
CMS <sub>O</sub> ; mediana (RI)				
≤30	14-24	23 (15, 38)	10 (5, 22)	15 (10, 25)
31-60	57-83	45 (25, 55)	22 (13, 28)	20 (15, 30)
≥61	32-39	63 (48, 79)	34 (27, 39)	30 (20, 45)
<i>p</i> -valor		<0.0001	<0.0001	<0.0001
PCS; media (DE)				
<50	91-134	42 (20)	21 (11)	22 (12)
≥50	14-14	76 (16)	36 (10)	39 (12)
<i>p</i> -valor		<0.0001	<0.0001	<0.0001
Barthel; mediana (RI)				
	<b>n</b>			
80-90	11-15	18 (12, 57)	12 (3, 27)	15 (10,30)
95	12-15	39 (22, 60)	23 (10, 27)	15 (15,30)
100	60-79	50 (42, 68)	27 (18, 33)	25 (15,40)
<i>p</i> -valor		0.004	0.003	0.019

AVD: actividades de la vida diaria. La puntuación total de ASES-p oscila entre 0-100 puntos. El rango de los valores de las sub-escalas de función y dolor es de 0 a 50 puntos cada una. CMS<sub>O</sub>: Puntuación Constant original; CMS<sub>R</sub>: Puntuación Constant relativa, estandarizada según edad y sexo; CMS<sub>NS</sub>: CMS sin el componente de la fuerza, con rango de valores 0-75. Los valores de los 4 componentes CMS<sub>O</sub> se derivaron de CMS original; PCS: puntuación del componente físico y MCS: puntuación del componente mental de SF-36. Grupos conocidos (n): representa el mínimo y máximo número de sujetos en las tres puntuaciones presentadas. Todos los análisis se basaron en los datos disponibles, no se realizaron imputaciones.

La sub-escala de la función presentó casi el mismo patrón de correlaciones, aunque con valores más bajos. Cabe mencionar que entre las correlaciones estudiadas de esta sub-

escala y los cuatro componentes de CMS<sub>0</sub>, la más elevada fue con AVD. Finalmente, la sub-escala del dolor presentó correlaciones más elevadas con el componente del dolor de CMS<sub>0</sub>, la dimensión del dolor y el PCS de SF-36. El resto de las correlaciones fueron más bajas. En relación a los grupos conocidos, sujetos con mejor estado de salud tenían valores ASES-p más elevado, aunque la sub-escala del dolor no diferenció bien entre los 3 grupos Barthel.

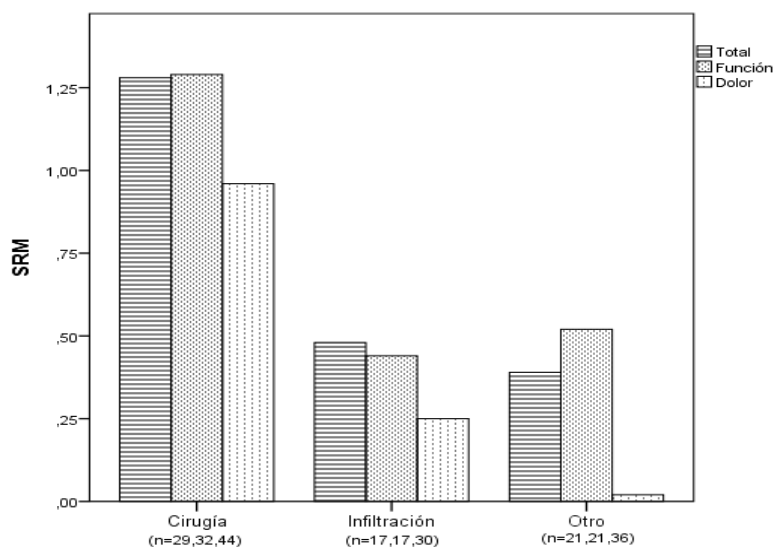
### ***Sensibilidad al cambio***

Durante el seguimiento, se obtuvieron n=120 respuestas validas. De ellas, n=10 no tenían contestado el ítem del dolor y n=50 dejaron en blanco por lo menos uno de los ítems de la sub-escala de la función, de la escala de interés. Tras estudiar las respuestas validas se observó que las diferencias, en las puntuaciones de la escala y los efectos estudiados, fueron más elevadas en los sujetos cuyo estado había mejorado. La puntuación total y la puntuación de la función presentaron valores de efectos alrededor de 1, mientras que para la sub-escala del dolor los efectos fueron de moderados a bajos (Tabla 6.5). Las correlaciones entre los valores pre y post de ASES-p oscilaron de 0.230 hasta 0.408, por lo que las estimaciones SRM<sub>Adj</sub> fueron casi idénticas con las estimaciones SRM (no se presentan resultados). Finalmente, los efectos, según el tipo de la intervención, fueron más elevados en los pacientes intervenidos (Figura 6.2).

**Tabla 6.5:** Sensibilidad al cambio. Puntuaciones ASES-p del hombro afectado, en el momento basal y al finalizar el seguimiento. La información se presenta de manera global y por estado de mejora

ASES-p	n	Basal media (DE)	Diferencia media (DE)	p-valor	SES	SRM
<b>Puntuación total (0-100 puntos)</b>						
Todas las respuestas	67	46.4 (23.5)	18.9 (25.2)	<0.001	0.80	0.75
Mejora	41	50.4 (23.9)	24.4 (23.8)	<0.001	<b>1.02</b>	<b>1.03</b>
No mejora	26	40.0 (21.9)	10.1 (25.3)	0.052	0.46	0.40
<b>Función (0-50 puntos)</b>						
Todas las respuestas	70	22.1 (11.8)	10.8 (13.4)	<0.001	<b>0.91</b>	<b>0.80</b>
Mejora	42	23.5 (11.8)	13.5 (11.6)	<0.001	<b>1.15</b>	<b>1.17</b>
No mejora	28	20.0 (11.8)	3.8 (14.0)	0.174	0.33	0.27
<b>Dolor (0-50 puntos)</b>						
Todas las respuestas	110	20.1 (12.9)	6.6 (14.9)	<0.001	0.50	0.45
Mejora	55	27.7 (13.0)	9.5 (14.6)	<0.001	0.73	0.65
No mejora	55	20.5 (12.3)	3.7 (14.8)	0.068	0.29	0.25

Las diferencias se derivaron como valores post tratamiento – valores basales; por lo tanto, las diferencias positivas indican mejor estado de salud. DE: desviación estándar. Los p-valores se basan en el t-test para muestras emparejadas. SES: tamaño de efecto estandarizado. SRM: respuesta media estandarizada.



**Figura 6.2:** Respuesta media estandarizada de la puntuación total y de las sub-escalas de ASES-p, según el tratamiento recibido. SRM: Respuesta media estandarizada. Los valores que se presentan debajo de cada grupo de tratamiento (n=...) indican la frecuencia de las respuestas validas para las puntuaciones total, función y dolor respectivamente.



## 6.4 Discusión

El proyecto actual, proporciona datos sobre la validación de la escala auto-administrada ASES-p en castellano. La fiabilidad, la validez y la sensibilidad al cambio se estudiaron considerando conjuntamente los 11 ítems de la escala y los 10 ítems de la sub-escala de la función, en diferentes modelos. Los valores de Cronbach fueron elevados y en concordancia con resultados presentados en validaciones previas (Michener et al. 2002, Yahia et al. 2011, Goldhahn et al. 2008, Piitulainen et al. 2014, Padua et al. 2010, Knaut et al. 2010, Celik et al. 2013). En general los resultados de los modelos AFC y Rasch, fueron satisfactorios.

En relación a los modelos de AFC, las cargas factoriales y la mayoría de los índices fueron aceptables y dentro de los límites deseados, en los diferentes modelos y solo el valor RMSEA para el modelo de los 10 ítems, se mostraba al margen de lo aceptable. El modelo Rasch, apoyó la unidimensionalidad de la escala y los resultados obtenidos no sugieren que existan ítems no productivos en la misma.

Aun así, los presentes análisis dirigen la atención hacia el ítem 10. “realizar su deporte habitual”. Este ítem en particular, presentó el mayor número de respuestas perdidas en el momento basal; tenía correlaciones bajas con otros dos ítems de la escala; la correlación más baja ítem-total; y sus índices de ajuste fueron más elevados que lo deseable (Bond & Fox 2007). Dado que este cuestionario ASES-p se desarrollo en Estados Unidos, la pregunta del deporte habitual puede reflejar una importante AVD de las personas en aquel ámbito, pero no necesariamente en el nuestro o por lo menos no en todos los grupos poblacionales. Información relacionada con los hábitos de la actividad física en nuestro entorno, sugiere que los sujetos >65 años no suelen realizar este tipo de actividad de manera regular (Departamento de Salud del Gobierno Vasco 2007). La edad media de los que no respondieron a este ítem, está de acuerdo con esta observación. Por otro lado, también cabe destacar que las patologías del hombro no impiden necesariamente la actividad física ya que varios tipos de deporte pueden ser compatibles con ellas (Clifford & Mallon 2005). Esta podría ser otra posible explicación

de la baja correlación entre el dolor y el ítem 10 en nuestra muestra y, consecuentemente, la falta de un mejor ajuste de este último ítem al modelo Rasch. En los AFC el ítem 10 obtuvo las cargas factoriales más bajas, pero en todo caso fueron  $>0.50$ . En general, este ítem no se consideró degradante para la escala y los datos actuales no justificarían se exclusión del instrumento de medición (Wright & Linacre 1994).

Otra observación interesante fue el nivel de *dificultad* (rank) del ítem del dolor, que resultó ser el cuarto más fácil para los participantes. Se esperaría que el dolor presentara una mayor dificultad, especialmente considerando que este ítem constituye el 50% de la puntuación total de ASES-p. Sin embargo, los niveles basales de dolor experimentados por los sujetos del estudio, fueron efectivamente moderados. Cabe además mencionar que este ítem evalúa la intensidad del dolor en el momento de la cumplimentación de la escala. La intensidad actual puede no representar el dolor general de la persona y no reflejar bien el estado de su patología. Esto podría explicar por qué sujetos con mejor función hayan presentado niveles más altos de dolor y viceversa. Es posible que esta característica en particular sea responsable del elevado outfit del ítem del dolor, ya que este valor sugiere la presencia de outliers (Bond & Fox 2007).

Asimismo, los resultados obtenidos sugieren que los ítems 7. "levantar 4.5kg por encima del hombro" y el ítem 8. "lanzar una pelota por encima de la cabeza" eran los más difíciles en realizar, mientras que el ítem 4. "limpiarse tras la micción y defecación" fue el más fácil. Estos resultados tienen sentido clínico. Las primeras dos acciones imponen una elevada tensión a la articulación, especialmente a los mecanismos que controlan su estabilidad. Además, se conoce que la repetición frecuente de estas acciones aumenta la prevalencia de las patologías del hombro (Reinold & Gill 2010). Por otro lado, las rotaciones internas, que se requieren por el musculo subescapular a la hora de limpiarse tras la micción y defecación, son más fáciles de realizar, incluso con el lado afectado (Escamilla et al. 2009). Artículos anteriores indican los ítems 7 y 4 como el más difícil y el más fácil en realizar (media más baja y alta respectivamente) (Goldhahn et al. 2008). El ítem 4 también presentó el valor medio más elevado en otra validación cultural, aunque

en aquel caso el ítem 3. “lavarse la espalda” resultó ser el más difícil (Piitulainen et al. 2014).

Sin embargo, las cargas factoriales y los resultados Rasch presentados en este estudio no son directamente comparables a los resultados de publicaciones anteriores. Otros autores han implementado análisis factorial exploratorio y de componentes principales (Yahia et al. 2011) y han sugerido uno (Piitulainen et al. 2014) o dos factores latentes (Yahia et al. 2011) para los ítems de la función.

En cuanto a la validez convergente, discriminante y de grupos conocidos, la puntuación ASES-p total presentó correlaciones más altas con todas las versiones de CMS, con los componentes de dolor y AVD de CMS<sub>0</sub> y con las dimensiones físicas y el PCS de SF-36. Correlaciones más bajas se encontraron con el RM y la fuerza de CMS<sub>0</sub>, así como con las dimensiones y componente mentales de SF-36. La sub-escala de la función presentó resultados similares. La sub-escala del dolor presentó correlaciones más altas con el componente CMS<sub>0</sub> de dolor, la dimensión de dolor corporal y el PCS de SF-36. Todas las correlaciones con el índice de Barthel fueron más bajas. Dado que el 50% de la puntuación total de ASES-p corresponde a la dificultad en realizar una serie de AVD, se consideró que las dos escalas presentarían correlaciones positivas y considerables. Sin embargo, a pesar de tener ciertas dificultades en sus AVD, los participantes de la presente muestra eran independientes en cuanto las AVD básicas. Este hecho explica el pequeño rango de los valores Barthel obtenidos y, como consecuencia, la baja correlación con ASES-p. Los resultados de los grupos conocidos también fueron satisfactorios, aunque la sub-escala del dolor no diferenció bien entre las primeras dos categorías de Barthel. Este hecho no sorprende ya que el Barthel no evalúa la presencia de dolor per se (Mahoney & Barthel 1965).

En las validaciones culturales de ASES-p, publicadas anteriormente, la validez convergente y divergente se estudió con las escalas Penn (Michener et al. 2002), DASH o SPADI (Yahia et al. 2011, Goldhahn et al. 2008, Celik et al. 2013), SST (Piitulainen et al. 2014) y OSQ (Padua et al. 2010), mientras que solo Yahia et al. no implemento ninguna versión de SF-36. Las correlaciones entre ASES-p y SF-36, en nuestro proyecto, eran

similares a las publicadas por Goldhahn (Goldhahn et al. 2008) y generalmente más altas de las presentadas en otros artículos de validación (Piitulainen et al. 2014).

Los presentes datos apoyan la sensibilidad al cambio de la escala. Los efectos estandarizados fueron moderados para la puntuación total de la escala, pero aquellos pacientes cuyo estado había mejorado al final del seguimiento, tenían efectos más elevados en comparación con el resto. Estimaciones similares se obtuvieron para la sub-escala de la función. Cuando la sensibilidad al cambio se estudió además según el tipo del tratamiento recibido, los pacientes que se sometieron a cirugía tenían efectos más elevados que el resto. La sub-escala del dolor presentó en todo momento, una sensibilidad al cambio baja o moderada.

El proyecto actual tiene ciertas limitaciones. Nuestro estudio se basa en una muestra de pacientes reclutados en centros hospitalarios públicos, por lo que los resultados aquí obtenidos puede que no sean aplicables a pacientes procedentes de otros ámbitos. También es posible que estos resultados no se puedan extrapolar a sujetos con fracturas de hombro. Dado que la mayor parte de la información se recogió por un cuaderno de recogida de datos auto-administrado, no disponemos de respuestas completas para todos los participantes. En el presente proyecto no se imputó ningún valor perdido en la escala ASES-p, todos los análisis se basaron en las respuestas proporcionadas por los participantes. En relación a la administración de la escala funcional de CMS cabe mencionar que existió cierta variabilidad, en cuanto la medición del componente de la fuerza. Según los recursos de cada centro participante, se implementaron diferentes tipos de pesas (mancuernas y pesas regulables). Por este motivo, los componentes de CMS se consideraron además por separado y de manera adicional también se estudió el CMS sin el componente de la fuerza.

Hasta donde sepamos, este estudio es la primera validación cultural de ASES-p que implementa AFC, así como el modelo Rasch. Asimismo, es el primero en considerar el ítem del dolor como parte de estos análisis. Futuros proyectos podrían enfocarse en la validez de constructo de la escala en diferentes ámbitos o patologías concretas, mientras que la asignación de puntuaciones en los diferentes componentes de la escala,

puede ser otro campo de estudio. Los presentes resultados, basados en la teoría clásica y en el modelo Rasch, sugieren que la versión castellana de ASES-p es válida y fiable para la evaluación de la patología de hombro en nuestro entorno. La dificultad en realizar el deporte habitual no fue un ítem informativo pero tampoco degradante para la escala. Los datos actuales apoyan la unidimensionalidad de ASES-p.



## **7. Fiabilidad test-retest del cuestionario ASES-p**

Este capítulo presenta los datos sobre la fiabilidad test-retest de la escala ASES-p. Dichos datos se recopilieron mediante entrevistas telefónicas.

## 7.1 Introducción

Inicialmente, los datos para estudiar la fiabilidad test-retest del cuestionario ASES-p se recogieron junto con los datos de la validación, presentados en el capítulo 6. En el momento basal de aquel estudio, a los primeros 130 participantes se les proporcionó de manera adicional, aparte de las correspondientes preguntas iniciales, el cuestionario ASES-p y una pregunta de cambio. Se les indicó el margen del tiempo en el cual tenían que contestar, solamente a estas preguntas, después de rellenar la batería de preguntas basales y se realizó un seguimiento telefónico. A pesar de las instrucciones proporcionadas y el seguimiento realizado no se consiguió recoger la información requerida dentro del plazo establecido, que era un máximo de 10 días.

En total se recibieron n=68 cuestionarios retest con fecha de cumplimentación de 5 a 66 días después del cuestionario basal. Las puntuaciones proporcionadas por n=58 sujetos estables, según la pregunta del cambio, indicaron una variabilidad muy amplia entre las dos administraciones consecutivas de la escala. La concordancia entre ítems fue reducida, las diferencias en las puntuaciones de las dos sub-escalas oscilaban entre +/- 20 puntos y la variabilidad para la puntuación total fue de +/-30 puntos. Además, limitar el análisis de fiabilidad test-retest a los sujetos que cumplían el plazo del tiempo pre-establecido fue imposible, debido a la reducida frecuencia de los mismos (n=9). Por este motivo, los datos requeridos para estudiar la validez test-retest se volvieron a recopilar de nuevo, esta vez vía entrevistas telefónicas. Para evitar sobrecargar a los participantes y minimizar las faltas de seguimiento, en esta fase se administró solamente la escala de ASES-p, acompañada por una pregunta de cambio en la segunda llamada.

## 7.2 Metodología

Se contactaron por teléfono todos los sujetos que proporcionaron respuestas válidas en el estudio presentado en el capítulo anterior. A los que aceptaron participar se les administró dos veces el cuestionario ASES-p. La segunda administración se realizó en un periodo de 3-7 días después de la primera. Durante la misma, se les preguntó a los participantes si habían experimentado algún cambio de salud desde la llamada anterior.



Los sujetos con respuestas positivas fueron excluidos de los análisis aquí presentados. Todas las entrevistas se realizaron por una persona experta en este tipo de recogida de datos, familiarizada con el cuestionario de interés. Durante la realización del retest, la persona que realizó las entrevistas, estaba ciega a las respuestas obtenidas en la primera llamada. Se estudió la concordancia exacta, de todos los ítems, en términos de porcentajes de respuestas idénticas. La concordancia en las dimensiones de dolor y función, así como de la puntuación total, se estudió principalmente mediante los límites de concordancia (limits of agreement) al 95% (LC 95%) de Bland y Altman (Bland & Altman 1986). Este método presenta una estimación de la variación, que se puede esperar, entre dos administraciones de la misma escala. Se basa en las diferencias de las puntuaciones pre-post y la desviación estándar de las mismas. Los LC 95%, son una función de estas estimaciones. El método de Bland & Altman se utiliza para estudiar la variabilidad de las respuestas que se obtienen entre diferentes momentos o evaluadores, bajo condiciones prácticamente iguales. La aceptabilidad de los correspondientes límites se tiene que evaluar según las características y los objetivos de cada instrumento de medición. La representación gráfica de estos resultados se realiza con diagramas de dispersión, en los cuales se dibujan las diferencias y los LC al 95% (y-axis) frente a las medias de los valores pre-post (x-axis). Aparte de eso, también se derivaron las concordancias exactas y los CCI de las puntuaciones. Los datos categóricos se presentan con frecuencias y porcentajes (%), las variables continuas con medias y desviaciones estándar (DE). Las comparaciones de medias se realizaron con el t-test de Student. Las variables categóricas se compararon con el chi-cuadrado. En este estudio no se realizaron imputaciones de datos perdidos.

### **7.3 Resultados**

Entre abril y septiembre de 2016 se contactaron todos los n=161 sujetos descritos anteriormente (capítulo 6). No hubo ningún fallecido, mientras que n=27 no aceptaron participar en esta fase adicional del proyecto. Por lo tanto, en el momento basal se obtuvieron un total de n=134 respuestas. Solo el ítem 10 “realizar su deporte habitual” presentó respuestas perdidas, ya que fue contestado por n=73 sujetos. El resto comentó

que actualmente no realizaba ningún deporte o que su única actividad física eran los paseos ligeros. En el mismo grupo de sujetos n=104 habían contestaron a esta pregunta en la recogida de datos anterior. La edad media de los participantes fue de 63.5 (DE: 11.6) años y 51% fueron mujeres. Los que no realizaban deporte eran 7 años mayores que el resto (67.5 vs. 60.1 años; p=0.0001), mientras que no se observaron diferencias de género (p=0.220) entre este grupo y el resto de la muestra. En la segunda llamada se obtuvieron n=94 entrevistas. El estado de salud de n=12 sujetos había cambiado desde la primera llamada, por lo que estos se excluyeron del estudio. Los análisis test-retest que aquí se presentan se basan en un máximo de n=82 sujetos estables. El ítem 10 obtuvo finalmente n=46 respuestas válidas para los análisis test-retest. La puntuación de la sub-escala del dolor se basa en todos los sujetos estables, mientras que debido a las pérdidas del ítem 10, la puntuaciones de la función y de la escala total se basan en n=46 sujetos.

**Tabla 7.1:** Rango de respuestas basales y concordancia de ítems y puntuaciones de la escala ASES-p

Ítem no.	Descripción	Min (%)	Max (%)	Concordancia exacta %
Dolor				
ítem	Intensidad de dolor hoy	0 (68)	10 (1)	73
Función				
ítem 1	Ponerse un abrigo	1 (7)	3 (73)	83
ítem 2	Dormir sobre lado afectado	0 (11)	3 (61)	80
ítem 3	Lavarse la espalda	0 (7)	3 (65)	78
ítem 4	Limpiarse tras micción/defecación	1 (10)	3 (78)	82
ítem 5	Peinarse	0 (4)	3 (74)	88
ítem 6	Llegar a una balda alta	0 (6)	3 (54)	77
ítem 7	Levantar 4.5kg por encima del hombro	0 (16)	3 (46)	67
ítem 8	Lanzar pelota por encima de la cabeza	0 (12)	3 (44)	68
ítem 9	Realizar trabajo habitual	0 (5)	3 (66)	89
ítem 10‡	Realizar deporte habitual	0 (41)	3 (50)	87
<b>Puntuación</b>				
ASES-p total (n=46)		22.5 (2)	100 (28)	33
Dolor (n=82)		0 (1)	50 (68)	73
Función (n=46)		5 (2)	50 (28)	39

El rango de las respuestas del ítem de dolor es de 0: ningún dolor a 10: intensidad máxima. El rango del resto de los ítems es de 0: imposible a 3: sin ninguna dificultad. Min (%) y Max (%): indican el mínimo y máximo valor obtenido en cada ítem de la escala en el momento basal. Los respectivos % indican el porcentaje de estos valores en el total de las respuestas. El ítem 10 se basa en n=46 respuestas, el resto de los ítems se basa en n=82 respuestas test-retest. Concordancia exacta %: se refiere al porcentaje de las diferencias iguales a 0 entre los dos momentos.

Los mínimos y máximos valores obtenidos en el momento basal, junto a sus respectivos porcentajes, la concordancia exacta, de los ítems y las puntuaciones de ASES-p, se presentan en la Tabla 7.1. En su mayoría, los participantes no experimentaban ningún dolor en la articulación del hombro, ya que el 68% de ellos asignó un valor de 0 a esta pregunta. Además, la mayor parte de la muestra (>50%) podía realizar casi todas las actividades de la escala “sin ninguna dificultad”, mientras que esta respuesta fue elegida por el 44-46% en los ítems 7.levantar 4.5kg por encima del hombro y 8.lanzar una pelota por encima de la cabeza. El único ítem que resultó ser imposible a realizar para el 41% de los participantes fue el ítem 10.realizar su deporte habitual. La puntuación total de ASES-p, derivada de las respuestas validas osciló entre 22.5 y 100 puntos, mientras que las dos sub-escalas recibieron de 0-50 (dolor) y de 5-50 (función) puntos.

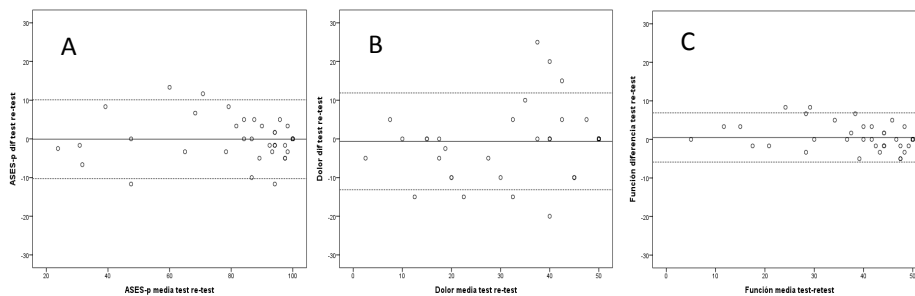
La concordancia exacta, en 9 de los 11 ítems, fue entre 73% y 89%, mientras que solo los ítems 7 y 8 presentaron valores de 67-68%. La puntuaciones total y de la función, presentaron diferencias test-retest igual a 0 en <40% de los participantes. Dado que la sub-escala del dolor se calcula directamente del respectivo ítem, la concordancia exacta de ambos fue la misma. Los CCI de las tres puntuaciones fueron los siguientes. Puntuación total: CCI=0.97 (IC 95%: 0.95-0.99); dolor: CCI=0.88 (0.83-0.93); función: CCI=0.96 (IC 95%: 0.93-0.98). Las medias diferencias en las puntuaciones test-retest fueron prácticamente 0 para la escala total y de -0.64 y 0.51 puntos para el dolor y la función (Tabla 7.2). El coeficiente de fiabilidad test-retest para toda la escala fue igual a 10.2 puntos. Los coeficientes del dolor y de la función fueron 12.5 y 6.4 respectivamente. Esto implica que la puntuación ASES-p total, entre las dos administraciones, varió  $\pm 10$  puntos. La variación de la sub-escala del dolor fue más amplia, mientras que la función tuvo una menor variación.

**Tabla 7.2:** Diferencias test-retest y límites de concordancia de las puntuaciones ASES-p

Puntuación	Rango	Diferencia	LC 95%
ASES-p total (n=46)	0-100	-0.09	-10.3 a 10.1
Dolor (n=82)	0-50	-0.64	-13.2 a 11.9
Función (n=46)	0-50	0.51	-5.9 a 6.9

Diferencia=valor pre menos valor post. Rango, es el posible rango de puntuaciones total y en las dos sub-escalas. Los límites de concordancia se estimaron con el método de Bland & Altman 1986. Frecuencia (n=) indica las respuestas validas en las que se basan las respectivas estimaciones.

Los límites de concordancia test-retest para las diferentes puntuaciones de ASES-p se presentan en la Figura 7.1. En el caso del dolor, los límites de concordancia sugieren, que la variación obtenida entre los dos momentos puede llegar hasta un 25% de la puntuación total de esta sub-escala (i.e. 12.5/50 puntos). Esta observación quizás contrasta con el elevado porcentaje de concordancias vistas en el dolor. Cabe aclarar que n=52 sujetos estaban libres de dolor en ambos momentos, pero los otros n=30 participantes tuvieron diferentes niveles de dolor. Tal y como se aprecia en la Figura 7.1, las diferencias test-retest de este último subgrupo podrían llegar hasta los 20 o más puntos.



**Figura 7.1:** Límites de concordancia al 95% para las puntuaciones total (A), dolor (B) y función(C), de la escala ASES-p.

## 7.4 Discusión

Los resultados obtenidos apoyan la fiabilidad test-retest de la puntuación total y de la sub-escala de la función, pero no la del dolor. La concordancia exacta en casi todos los ítems fue elevada, aunque cabe destacar el nivel de concordancia más bajo que se observó en los ítems 7 y 8. Ambos ítems tienen que ver con acciones que suponen un esfuerzo elevado para la articulación del hombro (Reinold & Gill 2010). Además, según los análisis presentados en el capítulo anterior, estas dos preguntas fueron precisamente las más difíciles para los participantes (Vrotsou et al. 2016). La menor concordancia observada en estos dos casos, se podría posiblemente atribuir al hecho de que los sujetos en realidad no realizan estas acciones. Durante la entrevista, varios

participantes comentaron que evitaban hacer las acciones que estos dos ítems proponían. A estos sujetos se les pedía contestar “*qué dificultad piensa que tendría en realizar esta acción, aunque no la realiza*”. Es más probable que varíen las respuestas de este tipo, sobre algo que no se realiza, frente a otras actividades de la vida diaria, como ponerse un abrigo o peinarse, por ejemplo.

Aun así, los límites de concordancia para la puntuación total de la escala se consideran aceptables e indican que los resultados de la escala son reproducibles, bajo condiciones similares. La sub-escala de la función presenta límites más estrechos y ajustados, mientras que la del dolor presenta mayor variación. Esta diferencia puede haberse producido por diferentes motivos. La puntuación de la función se basa en las respuestas de 10 ítems, mientras que la del dolor en una única pregunta. Además, debido a la particularidad de esta pregunta (nivel del dolor hoy), las respuestas pueden variar de un día al otro, incluso en las personas cuyo estado no haya cambiado. Finalmente, es también probable que el amplio rango de alternativas de respuesta de este ítem haya introducido variación adicional.

Nuestros resultados están de acuerdo con validaciones previas realizadas en otros idiomas (Michener et al. 2002, Goldhahn et al. 2008, Yahia et al. 2011, Celik et al. 2013, Padua et al. 2010, Kocher et al. 2005, Piitulainen et al. 2014). En todos estos estudios, las estimaciones CCI para la puntuación total fueron elevadas y  $>0.80$ . Mientras que la puntuación del dolor presentó CCI más bajos que la función y la puntuación total. En la validación Finlandesa, el dolor obtuvo un  $CCI < 0.70$  en sujetos estables, resultado que no se consideraría aceptable. Aunque es posible que esta discordancia se deba al largo periodo de 4 semanas entre test y retest de aquel estudio (Piitulainen et al. 2014).

Goldhahn et al. 2008, también presenta el porcentaje de respuestas idénticas entre las dos administraciones consecutivas de ASES-p. En aquella muestra, el ítem 8 fue el que menor concordancia tenía, mientras que el ítem 7 contaba con mayor porcentaje de coincidencias, que los ítems 3. lavarse la espalda y el ítem 10. Finalmente, solo Yahia et al. 2011 estudió la concordancia test-retest aplicando el método de Bland y Altman,

como nosotros. Los LC al 95% que estos autores dan para la puntuación total (i.e. -8.80 hasta + 10.16), se aproxima a nuestros datos.

Los datos aquí presentados dirigen otra vez la atención hacia el ítem del deporte, debido al elevado porcentaje de las no respuestas en el mismo. Sería interesante estudiar más al fondo este ítem y proporcionar quizás alternativas para sujetos con diferentes niveles de actividad física.

La limitación más importante de este proyecto, es el reducido número de pacientes en los que se basan las estimaciones de los límites de concordancia para las puntuaciones total y de la función. Es probable que las pérdidas observadas hayan introducido variación adicional en las estimaciones de estos límites, mientras que con una muestra más amplia los resultados podrían haber indicado una mayor reproducibilidad (Bland 2010). Se intentó obtener el máximo número de respuestas posible. Para no sobrecargar a los participantes se les administró solamente la escala de interés en ambos momentos, añadiendo la pregunta del cambio en la segunda llamada. Todas las entrevistas telefónicas se realizaron por una persona entrenada para tal fin y el seguimiento de todos los participantes fue muy estricto. Aun así las pérdidas de seguimiento fueron inevitables, ya que muchos participantes no quisieron volver a realizar la entrevista una segunda vez. Obviamente, las faltas de respuesta al ítem del deporte, resultaron en más datos perdidos. A pesar de esta limitación, se considera que tanto los límites de concordancia de la puntuación total, como los de la función, serían aceptables. Por otro lado, en su formato actual, la sub-escala del dolor no parece proporcionar puntuaciones reproducibles.

## **8. Cualidades psicométricas de la escala Constant-Murley Score**

Los datos presentados en este capítulo han sido enviados y están bajo evaluación, tras una revisión con cambios mayores, en la revista "Health Services Research": BHSR-D-16-01293: Constant-Murley Score: systematic review and standardised evaluation in different shoulder pathologies. Vrotsou K, Ávila M, Machón M, Mateo-Abad M, Pardo Y, Garin O, Zaror C, Gonzalez N, Escobar A, Cuéllar R.

## 8.1 Introducción

La escala de Constant-Murley Score (CMS) se presentó en 1987 como un instrumento apto para la evaluación de la funcionalidad del hombro (Constant & Murley 1987). Evalúa dos aspectos subjetivos, dolor y actividades de la vida diaria (AVD) y dos aspectos objetivos, rango de movimiento y fuerza. Las preguntas subjetivas se responden por el propio paciente, mientras que objetivas se basan en la observación clínica y se rellenan por el especialista. La escala asigna, respectivamente, 35 y 65 puntos a cada uno de sus dos componentes. La puntuación máxima total es la suma de ambos, es decir llega hasta un máximo de 100 puntos, valor que representa una articulación en óptimo estado funcional.

Como ya mencionamos en capítulos anteriores, este instrumento ha sido ampliamente utilizado para la medición y el seguimiento del hombro patológico (Barra-Lopez 2007, Kirkley et al. 2003, Angst et al. 2011). A pesar de ello, varias publicaciones subrayan una serie de aspectos problemáticos de esta escala. Se desconoce la metodología y los criterios implementados durante la selección de sus ítems, así como los motivos de la distribución de la puntuación total en los diferentes apartados del instrumento. Además, se cuestiona su aplicación a todas las patologías del hombro, mientras que también se ponen en duda los aspectos de fiabilidad y validez de la escala (Barra-Lopez 2007, Kirkley et al. 2003, Conboy et al. 1996, Othman & Taylor 2004, van den Ende et al. 1996). La falta de estandarización, a la hora de medir la fuerza de la articulación, y las diferencias observadas según la edad y el sexo de los sujetos han sido también comentadas (Lillkrona 2008, Bankes et al. 1998, Walton et al. 2007, Yian et al. 2005).

Como respuesta a todas estas críticas, el autor original publicó en 2008 un artículo con modificaciones e instrucciones específicas sobre el uso de CMS (Constant et al. 2008). En este trabajo, se aclaraba que el CMS no era apto para la evaluación del dolor episódico agudo, como sucede en los casos de luxación. El autor original propuso además una escala visual analógica (EVA) para la evaluación del dolor y parte de las preguntas sobre AVD y proporcionó instrucciones sobre cómo se debería evaluar el ítem de la fuerza.



Finalmente, en el mismo artículo, también se propuso una modificación de la puntuación total, ajustada por la edad y el sexo de los sujetos (Constant et al. 2008).

A lo largo de los años, las propiedades psicométricas de CMS han sido el objetivo de revisiones no sistemáticas (Kirkley et al. 2003, Angst et al. 2011), de revisiones sistemáticas sobre todo tipo de patologías del hombro (Roy et al. 2010) y de revisiones sobre patologías concretas (Slobogean & Slobogean 2011, Huang et al. 2015). Aun así, hasta la fecha, no existe una evaluación estandarizada que permita una percepción clara sobre las propiedades de la escala en las distintas patologías de hombro. El objetivo del trabajo actual fue realizar una revisión sistemática y evaluación estandarizada de las propiedades de CMS, en diferentes patologías de la articulación. En base a la información encontrada, se evaluaron las siguientes patologías: el síndrome subacromial, las fracturas, la artritis, la inestabilidad y el hombro congelado. Aquellos artículos con información sobre varias patologías presentadas conjuntamente y artículos en sujetos sanos se evaluaron por separado. Todas las evaluaciones se realizaron por revisores expertos, aplicando la herramienta EMPRO (Valderas et al. 2008) . El presente proyecto ofrecerá a los clínicos e investigadores información actualizada y una visión global de las propiedades métricas de CMS en diferentes patologías de hombro.

## 8.2 Material y Métodos

### ***Búsquedas sistemáticas***

Se realizaron búsquedas sistemáticas en las bases de PubMed e Embase durante el periodo comprendido entre 1 de enero 1986 y el 2 de mayo 2014. Los términos implementados en cada una de las dos bases fueron los siguientes.

En PubMed:

1. ((((((constant[All Fields] AND score[All Fields]) OR (constant[All Fields] AND murley[All Fields])) OR (constant[All Fields] AND murley[All Fields] AND score[All Fields])) OR (constant[All Fields] AND ("shoulder"[MeSH Terms] OR "shoulder"[All Fields]))) OR (constant[All Fields] AND ("shoulder"[MeSH Terms] OR "shoulder"[All Fields] AND score[All Fields])) OR constant-murley[All Fields]) OR (constant[All Fields] AND ("weights and measures"[MeSH Terms] OR ("weights"[All Fields] AND

*"measures"[All Fields]) OR "weights and measures"[All Fields] OR "scale"[All Fields])) AND ("1986/01/01"[PDAT] : "2014/05/02"[PDAT]).*

2. *("shoulder"[MeSH Terms] OR "shoulder"[All Fields]).*

3. *1 AND 2*

En Embase:

1. *'Constant'/exp OR 'constant murley (score)'/exp OR 'constant and murley score'/exp OR 'constant score'/exp ' OR constant murley score'/exp OR 'constant shoulder'/exp OR 'constant shoulder score'/exp OR 'constant-murley'/exp OR 'constant-scale'/exp OR constant shoulder score/de OR constant shoulder/de OR constant murley score/de OR constant murley/de OR constant score/de OR constant scale/de.*

2. *should\* OR 'shoulder'/exp OR shoulder.*

3. *3. 1 AND 2.*

4. *#3 AND [embase]/lim AND [1986-2014]/py*

Se incluyeron todos los artículos con información sobre el proceso de desarrollo del instrumento, las propiedades psicométricas y modos de administración del mismo. En la fase de la evaluación se incluyeron artículos escritos en inglés, español, francés, alemán e italiano. Se excluyeron, cartas de opinión, abstract y comunicaciones a congresos, protocolos de estudios, estudios de casos, estudios en animales o estudios en cadáveres. También se excluyeron artículos sobre técnicas quirúrgicas u otras relacionadas con el tratamiento de la articulación del hombro.

Los títulos, abstract y textos de los artículos seleccionados se revisaron en tres fases, de manera independientes por la autora de esta tesis (KV) y una de la coautoras del presente artículo (MA). Otra investigadora (YP) fue designada para resolver posibles discrepancias en caso de que se produjeran. Para completar la búsqueda, se realizaron además búsquedas manuales de las referencias bibliográficas de los artículos seleccionados. Aquellos artículos de revisión que no enfocaban a patologías concretas se revisaron solo por KV & MA, para identificar posibles referencias relevantes. Los artículos de revisión cuyo objetivo eran patologías concretas de hombro, se proporcionaron a los evaluadores. Aunque estos artículos no se evaluaban per se, su lectura podría aportar datos a considerar, mientras que también servían para la posible identificación de referencias adicionales.

Se anotaron las patologías de hombro de cada artículo identificado y a continuación se realizó una agrupación de las mismas según sus características. Los criterios de esta agrupación se establecieron considerando las indicaciones de la American Academy of Orthopaedic Surgeons (AAOS) y fue realizada por uno de los coautores (RC), especialista en cirugía ortopédica y con experiencia en la extremidad superior.

### ***Evaluaciones estandarizadas***

Los artículos de cada grupo de patología fueron evaluados, por dos revisores independientes, mediante la herramienta EMPPO. Los evaluadores eran expertos en el ámbito de las escalas de calidad de vida relacionada con la salud. La mayoría de ellos pertenecía al grupo que desarrollo EMPRO, o había realizado un curso de formación sobre el uso de esta herramienta. Una vez realizadas todas las evaluaciones EMPRO, cada evaluador pudo acceder a la evaluación de su “pareja”. Las parejas de evaluadores discutieron sobre las discrepancias ocurridas y llegaron a una evaluación conjunta y consensuada.

### ***Puntuaciones EMPRO***

Se evaluaron las siguientes propiedades psicométricas: *concepto y modelo de medida, fiabilidad (consistencia interna y reproducibilidad), validez, sensibilidad al cambio, interpretabilidad, facilidad de uso (carga para el entrevistado y carga de administración), y formas alternativas de administración* de CMS. A continuación se calcularon las puntuaciones EMPRO para cada grupo de patología, según lo indicado en la sección 4.4. El cálculo de dichas puntuaciones se realizó con el SPSS, versión 23 (SPSS, Chicago, IL, USA).

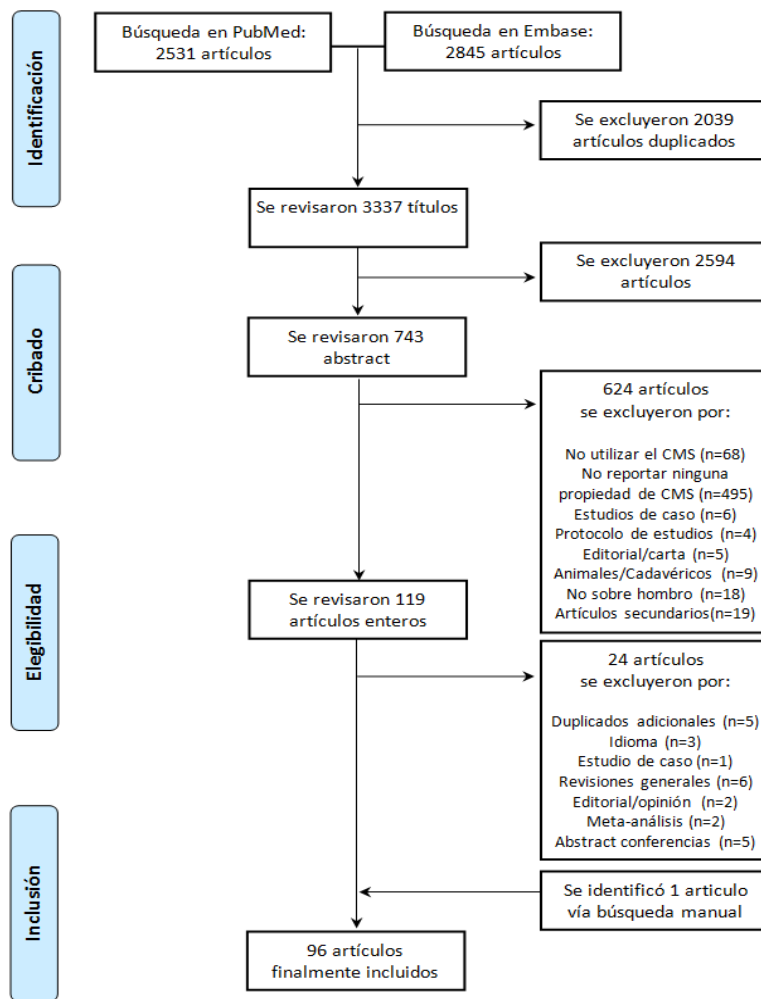
Dado que el CMS se desarrollo como una herramienta aplicable a todas las patologías de la articulación del hombro, la evaluación del atributo *concepto y modelo de medida* se tenía que basar en la misma información, en todos los grupos de patologías. Se consideró redundante pedir a todos los evaluadores revisar y puntuar la misma información, realizando evaluaciones repetidas de este atributo. En lugar de eso, dicho

atributo se evaluó solo una vez (por las autoras KV & MA) y la puntuación del mismo se incorporó en el cálculo de las puntuaciones totales de todos los grupos. Por otro lado, se consideró que la *carga de administración* y la *carga para el entrevistado* podrían depender de la patología estudiada. Por este motivo, ambos sub-atributos de la *facilidad de uso*, se evaluaron por separado en cada grupo. Lo mismo sucedió con las *formas alternativas de administración*. Ninguno de estos últimos atributos afecta la puntuación total de EMPRO.

### 8.3 Resultados

La búsqueda sistemática identificó un total de 3337 títulos únicos. De ellos, 2594 se excluyeron por no estar relacionados con el tema del estudio. De los 743 abstracts revisados 624 se excluyeron por no mencionar el instrumento CMS (68) o por no reportar ningún dato relacionado con las propiedades psicométricas de CMS (495). El resto se excluyó por ser investigación secundaria, estudios de casos, protocolos, comentarios o cartas de opinión, estudios cadavéricos, estudios en animales o estudios en patologías no relacionadas con el hombro. En la revisión de los textos enteros, se excluyeron 24 artículos adicionales, por no cumplir los criterios de inclusión y se identificó 1 durante la búsqueda manual de referencias. Por lo tanto, las evaluaciones EMPRO se basaron en un total de 96 artículos publicados (Figura 8.1).

Posteriormente, los artículos incluidos se dividieron en cinco grupos de patologías del hombro, concretamente, patología subacromial, fracturas, artritis, inestabilidad y hombro congelado. Estudios realizados con sujetos de varias patologías de la articulación y estudios en sujetos sanos, se evaluaron por separado. La información exacta sobre las patologías consideradas y el número de los artículos que se incluyeron en cada caso se presentan en la Tabla 8.1. Cada pareja de evaluadores revisó entre 1 (hombro congelado) y 37 (patología subacromial) artículos publicados. Aquellos artículos con información detallada sobre más de una patologías también fueron proporcionados a los evaluadores de las respectivas patologías. Los títulos de los artículos evaluados por cada grupo se presentan en el Anexo II.



**Figura 8.1:** Diagrama de flujo PRISMA con los pasos de la búsqueda bibliográfica.

Las puntuaciones EMPRO, de los diferentes grupos, oscilaron entre un máximo de 58.6 hasta un mínimo de 30.3 puntos, para los grupos de patologías subacromial y sujetos sanos, respectivamente. Debido a la falta de información, no se estimó una puntuación global para el grupo de hombro congelado. El grupo subacromial fue el único en recibir una puntuación total EMPRO >50 puntos. Los grupos de fracturas, artritis e inestabilidad recibieron 43.5, 41.7 y 30.6 puntos, mientras que el grupo de varias patologías obtuvo 49.3 puntos respectivamente (Tabla 8.2).

**Tabla 8.1:** Patologías incluidas en cada grupo de evaluación y número de correspondientes artículos evaluados

	Patologías incluidas	Nº artículos
<b>Patologías individuales</b>		
Patología subacromial	Síndrome de pinzamiento del hombro; manguito rotador; bursitis; tendinitis; tendinosis; tendinitis calcificada	37
Fracturas	Fracturas proximales del humero	7
Artritis	Osteoartritis glenohumeral; artritis reumatoide; enfermedad degenerativa de la articulación del hombro; necrosis avascular de la cabeza humeral	6
Inestabilidad	Inestabilidad traumática o no traumática; luxación recurrente; dislocación recurrente	5
Hombro congelado	Hombro congelado; capsulitis adhesiva	1
<b>Grupos adicionales</b>		
Varias patologías	Varias patologías; dolor de hombro	29
Sujetos sanos	Ninguna patología del hombro; individuos sanos	9
<b>Concepto y modelo de medida*</b>		2
<b>Total</b>		<b>96</b>

\*El atributo del concepto y modelo de medida se evaluó de manera global para la escala CMS.

El atributo de la *consistencia interna* se pudo puntuar solo para los grupos de patología subacromial y varias patologías. En ambos casos, los valores fueron bajos y no superaron los 38 puntos. Por otra parte, las puntuaciones recibidas por el atributo de la *reproducibilidad* fueron en general más altas. Los grupos de las patologías subacromial y de fracturas recibieron más de 50 puntos. Artritis e inestabilidad recibieron 33.3 y 0 puntos (mínimo posible), respectivamente, mientras que varias patologías, así como sujetos sanos obtuvieron puntuaciones >50. La falta de información basada en la teoría de respuesta al ítem penalizó las evaluaciones de la *reproducibilidad*. Las puntuaciones de *validez* en los 5 grupos de las patologías individuales oscilaron entre 44.7 y 31.3 puntos.

**Tabla 8.2:** Ítem, atributo y puntuación total de EMPRO en todos los grupos de evaluación

Atributo	Grupos de patologías individuales					Hombro Congelado	Grupos adicionales	
	Patología Subacromial	Fracturas	Artritis	Inestabilidad	Patologías Sanas		Varias Patologías	Sujetos Sanos
<b>Concepto y modelo de medida</b> †	33.3	33.3	33.3	33.3	33.3	33.3	33.3	33.3
<b>Fiabilidad: puntuación global</b>	70.8	58.3	33.3	0		33.3	54.2	62.5
Fiabilidad: consistencia interna	25						37.5	
Métodos de recolección descriptos	+++	-	-	-	-	-	+++	-
Cronbach/KR-20 adecuados	++	-	-	-	-	-	+++	-
Se proporcionan estimaciones TRI	-	NA	-	-	-	-	++	-
Pruebas en diferentes poblaciones	-	-	-	-	-	-	-	-
Fiabilidad: reproducibilidad	70.8	58.3	33.3	0		33.3	54.2	62.5
Métodos de recolección descriptos	++++	+++	++	+	-	-	+++	++++
Test-retest e intervalos adecuados	++++	+++	+++	+	-	-	+++	++++
Coefficientes de reproducibilidad adecuados	++++	+++	++	+	-	-	+++	++++
Se proporcionan estimaciones TRI	-	-	-	-	-	-	-	-
<b>Validez</b>	44.4	31.3	41.7	41.7			53.3	55.6
Validez de contenido adecuada	-	+	+	-	-	-	+	-
Validez de constructo/criterio adecuada	+++	+++	+++	++	-	-	++++	+++
Composición de la muestra descrita	+++	++	++++	++	-	-	++++	+++
Hipótesis previas descriptas	+++	+++	+++	+++	+	+	+++	+++
Justificación validez de criterio	NA	NA	NA	NA	NA	NA	NA	NA
Pruebas en diferentes poblaciones	-	-	-	+++	-	-	++	+++
<b>Sensibilidad al cambio</b>	83.3	50	55.6	50			55.6	-
Métodos adecuados	+++	+++	+++	+++	+	+	+++	-
Magnitud de cambio descrita	+++	+++	+++	+++	-	-	+++	-
Comparación de grupos estables e inestables	+++	++	++	+	-	-	++	-
<b>Interpretabilidad</b>	61.1	44.4	44.4	27.8			50	+++
Justificación de criterios externos	+++	+++	+++	++	-	-	+++	+++
Descripción de las estrategias de interpretación	+++	++	+++	++	-	-	++	-
Cómo presentar resultados descrito	++	++	-	++	-	-	+++	-
<b>Puntuación EMPRO total</b>	<b>58.6</b>	<b>43.5</b>	<b>41.7</b>	<b>30.6</b>			<b>49.3</b>	<b>30.3</b>
Facilidad de uso								
<b>Facilidad de uso I: para el entrevistado</b>	61.1		33.3	16.7			44.4	-
Habilidad y tiempo necesario	+++	-	+++	++	-	-	+++	-
Impacto en el entrevistado	++	+	++	++	-	-	+	++
Circunstancias no apropiadas	++++	-	-	+	-	-	+++	-
<b>Facilidad de uso II: carga de administración</b>	50	+++	83.3	20.8			75.0	-
Recursos necesarios	+	+++	++++	+	-	-	+++	-
Tiempo necesario	+++	-	+++	+	-	-	++++	-
Formación y experiencia necesarias	++	-	+++	+	-	-	+++	-
Carga asociada al cálculo de las puntuaciones	+++	-	+++	+++	-	-	+++	+
<b>Formas alternativas de administración</b>	+++	-	+++	+++			33.3	-
Características métricas y de uso	-	-	-	-	-	-	++	-
Comparación de las formas alternativas	-	-	-	-	-	-	++	-

Las puntuaciones de los ítems oscilan de muy de acuerdo (++++) a muy en desacuerdo (-); no aplicable (-); no información (-); no aplicable (NA). TRI: Teoría de respuesta a los ítems. La puntuación total EMPRO de todos los grupos incluye el concepto y modelo de medida (33.3 puntos). † Las puntuaciones de este atributo fueron: concepto de medida (++++); ítems y sus combinaciones (+); razonamiento sobre dimensionalidad (+); participación población diana (-); variabilidad (+++); nivel de medida (++) cálculo de puntuaciones (++)

Los grupos de varias patologías y sujetos sanos recibieron, en este atributo, 53.3 y 55.6 puntos respectivamente. La *sensibilidad al cambio*, con puntuaciones entre 83.3 y 50 puntos, fue el atributo mejor evaluado. Esta propiedad no suele ser relevante para los sujetos sanos ya que ningún cambio se espera, ni se experimenta, por lo menos a corto plazo, por estos sujetos.

En cuanto al atributo de la *interpretabilidad*, los grupos de la patología subacromial y de inestabilidad recibieron la más alta (61.1) y la más baja (27.8) puntuación respectivamente. Las otras dos patologías individuales obtuvieron 44.4 puntos cada una. El grupo de varias patologías consiguió 50 puntos, mientras que dicha puntuación no se calculó para los sujetos sanos. Los dos sub-atributos de la *facilidad de uso* del instrumento se puntuaron de manera muy variable entre los grupos considerados y solo el grupo de la patología subacromial obtuvo >50 puntos en ambos. Finalmente, el atributo *formas alternativas de administración* se evaluó solo para el grupo de varias patologías y recibió 33.3 puntos. Esta evaluación se basó en un único artículo que presentaba una versión auto-administrada de CMS. Una serie de instrucciones detalladas y fotografías explícitas guían a los sujetos sobre cómo realizar y auto-evaluar su rango de movimiento y la fuerza, inicialmente diseñadas para los especialistas (Levy et al. 2014). En el proyecto actual no se evaluó el atributo *proceso de adaptación cultural y lingüística*. La información basada en versiones adaptadas, de CMS en otros idiomas, se consideró parte de la evaluación estandarizada. Este enfoque se ha seguido en artículos anteriores (Schmidt et al. 2014).

## 8.4 Discusión

La escala CMS ha sido ampliamente aceptada e utilizada, sin haberse previamente sometido a un proceso de validación (Barra-Lopez 2007, Angst et al. 2011, Constant et al. 2008, Roy et al. 2010). En el presente estudio, las cualidades psicométricas de CMS se evaluaron en siete grupos de patologías de hombro, utilizando el instrumento EMPRO.

En general, las puntuaciones asignadas fueron bajas. Los grupos *subacromial* y de *varias patologías* recibieron mejores evaluaciones que los demás, pero solo el primero superó



el límite de los 50 puntos EMPRO. Los atributos evaluados en el grupo de los *sujetos sanos* recibieron en general puntuaciones más altas en comparación con otros grupos de patologías. Esto se podría atribuir al hecho que el objetivo de la mayoría de las publicaciones de este grupo era estudiar al menos una de las propiedades de CMS. Eran por lo tanto más probables en explorar, analizar y reportar adecuadamente la respectiva información. Los atributos no evaluados penalizaron la puntuación total de este grupo. En el grupo de las *fracturas*, solo la fiabilidad y la sensibilidad al cambio llegaron a una puntuación aceptable. Otros autores también opinan que la evidencia disponible no avala el amplio uso de CMS en este tipo de pacientes (Slobogean & Slobogean 2011, Baker et al. 2008). En el grupo *artritis*, la sensibilidad al cambio y la carga de administración fueron los únicos atributos con puntuación aceptable en el grupo. La aplicación de la escala CMS a sujetos con artritis reumatoide ha sido criticada, principalmente por la dificultad que tienen estos sujetos en realizar la prueba de la fuerza (Christie et al. 2009). Asimismo, en la *inestabilidad* solo la sensibilidad al cambio llegó a los 50 puntos. La *inestabilidad* del hombro, puede dar lugar a episodios de luxación y dolor agudo, empeorando la función general de la articulación. Sin embargo, estos síntomas suelen tener un carácter intermitente, motivo por el cual el CMS no es apto para esta patología (Conboy et al. 1996, Constant et al. 2008, Dawson et al. 1999, Plancher & Lipnick 2009). La información incluida en el único artículo sobre la patología de *hombro congelado*, no permitió la estimación de ningún atributo EMPRO (Othman & Taylor 2004), salvo el concepto y modelo de medida, común para todos los grupos. El objetivo del artículo de Othman 2004 fue, explorar los inconvenientes que la administración de CMS presenta en esta patología.

Instrumentos capaces de captar cambios producidos a lo largo del tiempo y libres de error aleatorio, son de gran importancia en el ámbito de la salud (Mc Dowell 2006, Ferguson 2009, Aaronson et al. 2002). En el presente proyecto, la *sensibilidad al cambio* fue la propiedad psicométrica mejor evaluada en todos los grupos, con la excepción del hombro congelado. Las evaluaciones apoyan la sensibilidad al cambio del instrumento CMS, especialmente cuando se utiliza en sujetos con patología subacromial. Sin embargo, hubiera sido deseable disponer de más información para el resto de los grupos. La *fiabilidad* fue el segundo atributo mejor valorado. Entre sus dos

componentes, la *reproducibilidad* era más adecuadamente reportada que la *consistencia interna*. La evidencia disponible parece apoyar la *reproducibilidad* de CMS en el caso de la patología subacromial y de los sujetos sanos. Los grupos de fracturas y de varias patologías también obtuvieron puntuaciones por encima de los 50 puntos, pero la información sobre todo en relación a las características y obtención de la muestra fue insuficiente. El grado en que la puntuación de una escala puede asignarse un significado fácilmente comprensible, es decir la *interpretabilidad*, fue aceptable para el grupo subacromial y al límite para varias patologías. Mientras que la *validez*, entendiendo como tal el grado en que un instrumento mide lo que pretende medir, fue aceptable solo en los grupos de varias patologías y sujetos sanos.

La evaluación del atributo del *concepto y modelo de medida* subraya que la información relacionada con la creación de CMS es escasa. Aunque el concepto que la escala pretende medir fue claramente explicado y la variabilidad de sus puntuaciones adecuadamente descritas, no existe información sobre la población diana. La justificación sobre los ítems incluidos y los componentes de la escala es insuficiente. Tampoco se justifica la distribución de las puntuaciones en los diferentes ítems y componentes, tal y como comentan autores anteriores (Barra-Lopez 2007, Kirkley et al. 2003, Angst et al. 2011). Finalmente, se observó una gran variabilidad en los atributos de la *facilidad de uso*. El grupo de la patología subacromial fue el único con una puntuación aceptable en la *carga para el entrevistado*. Este grupo, junto con los grupos de artritis y de otras patologías, obtuvieron las puntuaciones más elevadas en el sub-atributo de *carga de administración*.

Existe una revisión sistemática anterior, cuyo objetivo, al igual que el del proyecto actual, fue estudiar las propiedades psicométrica de la escala CMS (Roy et al. 2010). Basados en 34 artículos escritos en inglés y francés y realizando una síntesis narrativa, los autores concluyeron que la escala contaba con buena *validez convergente* y *fiabilidad*, y excepto la patología de la inestabilidad, el CMS tenía una excelente *sensibilidad al cambio*. Nuestros resultados apoyan también la *validez convergente* de la escala, salvo en el caso del hombro congelado. Sin embargo, están en desacuerdo con la generalización de las otras dos afirmaciones. Según las presentes evaluaciones, y

basados en una evidencia más amplia, la *fiabilidad* de CMS no se demuestra en los grupos de artritis inestabilidad y hombro congelado. A excepción de la patología subacromial, en el resto de grupos sería deseable disponer de más información.

La intervención administrada es un factor muy importante en la evolución de cualquier patología. Es por ello que nuestro grupo contemplo la posibilidad de realizar evaluaciones adicionales, re-agrupando los artículos finalmente seleccionados, esta vez en base a la intervención realizada. Sin embargo, realizar dicha re-agrupación resultó ser una tarea muy compleja. En general, en un único artículo se comentaba la realización de diferentes intervenciones para la misma patología, mientras que en muchos casos, la información relacionada con la intervención ni siquiera estaba disponible. Además en la mayoría de los casos, los resultados de los estudios no se presentaban por separado, según tipo de intervención, sino de manera global. Por estos motivos, las evaluaciones adicionales según tipo de intervención no pudieron ser realizadas.

Este trabajo cuenta con una serie de limitaciones. Después de las últimas recomendaciones realizadas por el propio autor de CMS (Constant et al. 2008), la versión modificada de la escala (i.e. EVA para el ítem del dolor y las AVD) o la puntuación ajustada según la edad y el sexo de los participantes han sido utilizadas en ciertas publicaciones. La mayoría de los artículos que presentaron estas puntuaciones “actualizadas”, también reportaron los valores “originales” de CMS. Cuando eso no era el caso, los valores actualizados se consideraron igual que los valores originales. La evaluación conjunta se justifica por el hecho que las modificaciones en la puntuación de una escala se perciben como mejoras de la propia escala y no como instrumentos diferentes. Además, debido al reducido número de estas publicaciones, no se podrían producir posibles evaluaciones adicionales. Una limitación adicional es el hecho de que la mayoría de los artículos no explican la forma exacta de evaluar el componente de la fuerza. Es posible que aquellos estudios que realizaron dichas mediciones con un dinamómetro digitalizado, u herramienta similar, tal y como recomienda el autor original (Constant et al. 2008) obtuvieran puntuaciones más precisas y reproducibles. Es importante señalar que la información que los diferentes autores proporcionan en relación a la administración de CMS, es un aspecto que tiene que ser mejorado.

Asimismo, la experiencia de los revisores puede haber introducido cierta variabilidad en los resultados obtenidos. Sin embargo, las instrucciones específicas de EMPRO y el consenso final de todas las evaluaciones, deberían haber minimizado el efecto de este sesgo. Además, el número de artículos identificados en cada grupo puede haber afectado las evaluaciones, aunque la revisión sistemática realizada debería haber minimizado la posibilidad de exclusión de artículos relevantes. Finalmente, la herramienta EMPRO se desarrolló para la evaluación de escalas auto-cumplimentadas por el paciente. Aun así, lo cierto es que instrumentos como el CMS también deberían tener las mismas propiedades métricas que las escalas auto-administradas, hecho que justifica los objetivos y el enfoque del presente estudio (Mc Dowell 2006, Mitchell et al. 2013).

Hasta donde sepamos, éste es el primer estudio en evaluar las propiedades psicométricas del instrumento CMS, en diferentes patologías de la articulación, tras realizar una revisión sistemática y evaluación estandarizada de la literatura. Además, es el primero en realizar dichas evaluaciones de un modo que permita comparaciones entre ellas. Nuestros resultados ofrecen una imagen más clara de las propiedades psicométricas de la escala CMS, indicando sus fortalezas y debilidades, mientras que están de acuerdo con múltiples observaciones y comentarios publicados anteriormente (Rocourt et al. 2008, Blonna et al. 2012, Slobogean & Slobogean 2011, Huang et al. 2015, Christie et al. 2009, Angst et al. 2008). Este trabajo debería ser de interés no solo para los clínicos, que administran la escala, sino también para aquellos investigadores que deseen obtener más datos sobre las propiedades de CMS. Estudiar dichas propiedades para diferentes tipos de intervención o desarrollar variaciones de la escala para diferentes patologías podrían ser posibles futuras líneas de investigación.

## 9. Discusión general

La presente tesis se ha desarrollado alrededor de la patología de hombro, las escalas que se utilizan para la evaluación de la misma y las propiedades psicométricas que estas escalas tienen. Concretamente, de interés fueron las escalas auto-cumplimentadas por los propios pacientes, con objetivo final seleccionar la más adecuada entre ellas, para su posterior adaptación y validación al castellano. También se estudiaron las características psicométricas de la escala Constant-Murley Score (CMS) que combina respuestas auto-cumplimentadas con observaciones del especialista. El CMS es un instrumento muy relevante en este ámbito, ya que muchos clínicos lo percibe como “regla de oro” en la evaluación de la articulación de hombro (Lillkrona 2008, Kirkley et al. 2003, Angst et al. 2011).

La evidencia disponible se localizó mediante revisiones sistemáticas en las bibliotecas de PubMed y Embase y la valoración de las propiedades pertinentes se realizó por revisores expertos. Todas las evaluaciones que se presentan en esta tesis, se realizaron aplicando la herramienta EMPRO (Valderas et al. 2008), cuyo desarrollo se basó en las directrices marcadas por el Medical Outcome Trust (Aaronson et al. 2002). Asimismo, todas las evaluaciones se realizaron por pares de revisores y fueron consensuadas entre ellos. La revisión de la literatura, la utilización de una herramienta estandarizada, así como la involucración de expertos, la evaluación por pares y el consenso final, son pasos

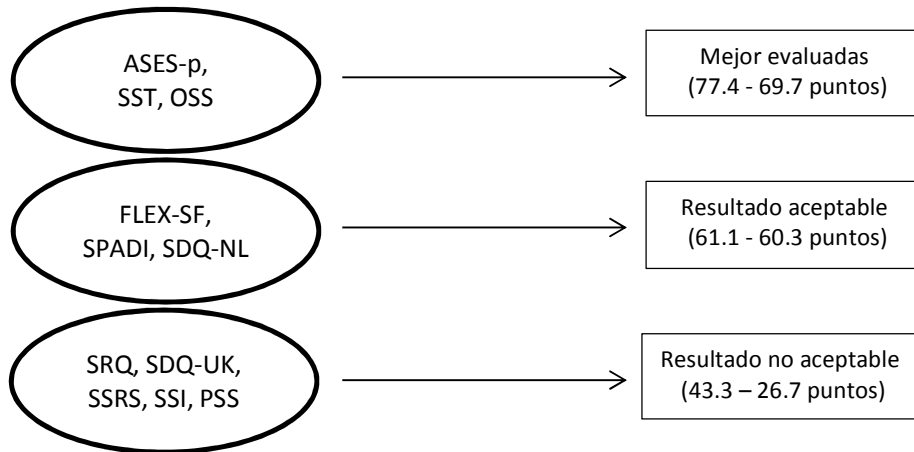
que garantizan un proceso riguroso y de calidad (Liberati et al. 2009). Los estudios que componen esta tesis ofrecen información actualizada y suponen un avance sobre lo previamente publicado por autores anteriores.

### ***En relación a las escalas auto-cumplimentadas evaluadas***

Debido a la aplicación de EMPRO, los resultados del primer estudio de esta tesis permiten no solo realizar una evaluación estandarizada de la evidencia, sino también una comparación directa de los datos obtenidos (Schmidt et al. 2014). Los resultados de este proyecto situaron las escalas *American Shoulder and Elbow Surgeons shoulder assessment, patient self-evaluation section (ASES-p)*, *Simple Shoulder Test (SST)* y *Oxford Shoulder Score (OSS)*, entre las tres mejor evaluadas. Mientras que también los cuestionarios *Flexilevel Scale of Shoulder Function (FLEX-SF)*, *Shoulder Pain and Disability Index (SPADI)* y el *Dutch Shoulder Disability Questionnaire (SDQ-NL)* sobrepasaron la puntuación total mínima aceptable. Al mismo tiempo, la evidencia disponible no apoyó el uso de los otros cinco cuestionarios evaluados: *Shoulder Rating Questionnaire (SRQ)*, *United Kingdom Shoulder Disability Questionnaire (SDQ-UK)*, *Subjective Shoulder Rating System (SSRS)*, *Shoulder Severity Index (SSI)* y *Penn Shoulder Score (PSS)*. La Figura 9.1, presenta un resumen de las puntuaciones EMPRO obtenidas.

Como en toda revisión sistemática, la evidencia encontrada condiciona los resultados obtenidos. Las evaluaciones de las cinco escalas peor evaluadas, se basaron de 2 (SSI) a 6 artículos (SRQ), mientras que las tres escalas mejor valoradas contaron con evidencia de 12-30 artículos cada una. Es posible que publicaciones no incluidas hayan perjudicado las evaluaciones. También es posible que el uso menos frecuente de ciertos instrumentos, haga que la información sobre sus características métricas no sea suficiente, para apoyar la calidad de las mismas. Nuestra búsqueda de la literatura y la posterior revisión manual de las referencias de todos los artículos incluidos, debería haber reducido el sesgo de omisión de artículos relevantes. Cabe mencionar que las cinco escalas peor evaluadas fueron creadas durante el mismo periodo de tiempo que la mayoría de los otros cuestionarios, por lo que la extensión de su uso no puede haberse

condicionado por este factor (Patte D. 1987, Croft et al. 1994, Kohn & Geyer 1997, L'Insalata et al. 1997, Leggin & Lannotti 1999).



**Figura 9.1:** Resumen de las evaluaciones EMPRO presentadas en el capítulo 5.

Otros estudios presentan datos comparables a los nuestros. En una evaluación anterior, los cuestionarios SRQ y SSI presentaron información aceptable solo en 3 de las 12 características evaluadas, los SDQ-UK y SSRS en 4, mientras que el PSS ni siquiera figuraba entre los evaluados (Bot et al. 2004). En relación al PSS, las publicaciones incluidas en nuestro trabajo indican que su uso se realiza, sobre todo, por miembros del equipo que creó la escala (Leggin et al. 2006, Leggin et al. 2003, Wylie et al. 2010, Michener et al. 2002, Cook et al. 2001). En otra revisión sistemática, cuyo objetivo era identificar los aspectos más frecuentemente utilizados en los instrumentos de hombro, el número de publicaciones de SRQ y PSS también fue reducido. En aquella revisión, limitada en artículos publicados entre enero 2005 hasta mayo 2010, no figuran las escalas SDQ-UK, SSRS y SSI (Roe et al. 2013).

Las escalas FLEX-SF, SPADI y SDQ-NL recibieron alrededor de 60 puntos EMPRO cada una, superando así el valor mínimo aceptable. El FLEX-SF se penalizó por falta de información relacionada con su *sensibilidad al cambio*. Esta escala, creada en 2003, es la más reciente de todas las evaluadas. Se trata de un cuestionario adaptable al nivel de la

función de hombro que presenta el paciente (bajo/medio/alto). Los sujetos contestan a una pregunta, que determina su nivel de función y, a continuación, rellenan una de las tres variaciones de FLEX-SF, la que corresponde a su nivel de función. La creación de este instrumento se basó en la metodología del modelo Rasch (Cook et al. 2003). Aunque información disponible, sobre las propiedades de FLEX-SF, se encontró solamente en 2 artículos, el contenido de los mismos fue suficiente para que dicho instrumento obtuviera la cuarta mejor evaluación. Es probable que datos adicionales proporcionen, en un futuro, aun mejor evidencia sobre las propiedades métricas de esta escala. Por otro lado, el cuestionario SPADI (Roach et al. 1991) está entre los más utilizados en el ámbito de la evaluación de hombro (Roe et al. 2013), hecho que se refleja en el elevado número de publicaciones, relacionadas con esta escala, identificadas en nuestra búsqueda (26). A pesar de ello, el atributo de la *interpretabilidad* de este instrumento no obtuvo una evaluación favorable. Según el artículo de Bot et al. 2004, el cuestionario SPADI cuenta con buena *interpretabilidad*, dato que contradice nuestra evaluación. Esta contradicción se puede explicar por las diferencias que presentan las dos herramientas de evaluación. No existe una correspondencia exacta de la información precisa que Bot et al. 2004 y EMPRO evalúan bajo atributos con nombres iguales. El número de preguntas por atributo, los matices de la información que cada herramienta pretender evaluar, así como el sistema de puntuación de las mismas (Bot et al. 2004-evaluación global por atributo: buena/dudable/mala/sin datos; EMPRO-varios ítems con 4 alternativas de respuesta y obtención de media), son diferencias que pueden llegar a producir discrepancias. Se entiende que una correspondencia exacta, entre las dos herramientas, incluso evaluando las mismas publicaciones, hubiera sido difícil.

Por último, según nuestra evaluación, el cuestionario SDQ-NL (van der Heijden et al. 2000) cuenta con muy buena *sensibilidad al cambio* y *validez*, pero su falta de información en *interpretabilidad* y la baja *fiabilidad* lo situaron en el sexto lugar. Esta escala, debido además a su excelente *facilidad de uso* (3 minutos), podría ser recomendable en proyectos longitudinal. Pero la *fiabilidad* de la misma, necesita estudiarse con más detalle. Bot et al. 2004 no encontró datos sobre la *fiabilidad* de este



instrumento. En 2005 un grupo de investigadores mejicanos realizó una adaptación de SDQ-NL al idioma diana (Alvarez-Nemegyei et al. 2005).

Entre las tres escalas mejor evaluadas, el ASES-p recibió puntuaciones más elevadas en los atributos de *concepto y modelo de medida* y *validez*. Por otro lado, el SST obtuvo la mayor puntuación posible en la *sensibilidad al cambio*. El OSS obtuvo mejores evaluaciones que el SST en los atributos de *concepto y modelo de medida* e *interpretabilidad*, pero en ningún caso estas puntuaciones fueron mayores que las de ASES-p. Según las evaluaciones obtenidas, cualquiera de estas tres escalas podría ser un buen candidato para traducir y validar al castellano. La decisión final, de seleccionar el ASES-p, se basó en las siguientes consideraciones. Por un lado, se tomó en cuenta la extensión de uso de esta escala (Longo et al. 2011, Kirkley et al. 2003, Slobogean & Slobogean 2011). Por otro, se consideró que las puntuaciones tipo Likert podrían ofrecer más flexibilidad, en comparación con las respuestas binarias de SST, en cuanto a la capacidad de discriminación de la escala, la concordancia test-retest, la sensibilidad al cambio y en general la recopilación de datos. Las respuestas tipo Likert suelen tener un mejor funcionamiento en este tipo de instrumentos (Preston & Colman 2000, Weng 2004). Finalmente, el hecho de que ASES-p exploraba el componente del dolor, como una dimensión separada con su propia puntuación, fue también un aspecto interesante. El dolor experimentado en la articulación de hombro es uno de los motivos de consulta más frecuentes en este ámbito (Singh et al. 2011, Bongers 2001, Luime et al. 2004, Marin-Gomez et al. 2006). Nos pareció por lo tanto relevante trabajar con una escala que exploraba este componente por separado.

En este punto, cabe señalar una observación general sobre los cuestionarios auto-cumplimentados, evaluados en nuestro estudio. Entre los cinco principales atributos, en los que se basa la puntuación total de EMPRO (i.e. *concepto y modelo de medida*, *fiabilidad*, *validez*, *sensibilidad al cambio*, e *interpretabilidad*), la *interpretabilidad* fue la característica peor valorada. La interpretación de las puntuaciones de una escala es un tema muy relevante tanto en la investigación como en la práctica clínica. Sin indicaciones claras, sobre valores de referencia de la población sana, de diferentes niveles de gravedad o valores mínimos clínicamente importantes, las escalas de

medición serían de utilidad reducida para los clínicos (Mc Dowell 2006, Guyatt 2000). En el ámbito de la medición de la articulación de hombro, la interpretabilidad todavía no se percibe como una propiedad tan importante como la *fiabilidad, la validez y la sensibilidad al cambio*. El trabajo de Bot et al. 2004, realizado una década antes que el nuestro, llegó a la misma conclusión sobre este atributo (Bot et al. 2004). Esto señala el considerable camino que todavía hace falta recorrer en esta dirección.

La evaluación de las escalas de hombro, realizada por nuestro grupo (Schmidt et al. 2014), ha tenido desde su publicación cierto impacto que cabe mencionar. Se cita en un artículo de validación, de la escala SST, al castellano (Membrilla-Mesa et al. 2015). También se referencia en un protocolo, publicado casi al mismo tiempo que el trabajo de Membrilla-Mesa 2015, sobre un estudio multi-céntrico con objetivo final la validación de la misma escala (Medina-Porqueres et al. 2015). Asimismo, un artículo de revisión a nivel internacional sitúa nuestra publicación entre el 20% de las revisiones realizadas entre 2013-2014, en escalas de medición, con descripción clara sobre el modo de síntesis de la evidencia (Terwee et al. 2016). De modo casi paralelo y completamente independiente de la evaluación aquí presentada, un grupo de investigadores de la Universidad de Alcalá de Henares, realizó la adaptación y validación castellana de los instrumentos OSS y SPADI, con mujeres que experimentaban problemas de hombro, tras una operación de cáncer de mama (Torres-Lacomba et al. 2015). Un artículo posterior, realizado por otro grupo investigador, presenta una validación más extensa de SPADI, con pacientes con patologías de hombro. En este último trabajo, sí que se citan los resultados obtenidos por Schmidt et al. 2014 (Luque-Suarez et al. 2016).

#### ***En relación al cuestionario ASES-p***

Los resultados presentados en el capítulo 6 de esta tesis, apoyan la unidimensionalidad de la versión castellana de ASES-p. Hasta donde sepamos, nuestro estudio es el primero en introducir el ítem del dolor en los modelos que exploran la estructura subyacente de la escala, y comprobar la existencia de un factor latente considerando todos los ítems de la misma (Michener et al. 2002, Yahia et al. 2011, Knaut et al. 2010, Celik et al. 2013, Padua et al. 2010, Piitulainen et al. 2014). Los resultados obtenidos, durante la fase de la

validación, apoyan la existencia de una puntuación global compuesta por los 11 ítems del cuestionario (Vrotsou et al. 2016).

La escala ASES se creó, principios de los años '90, cuando todavía los métodos matemáticos más avanzados, como son los modelos factoriales y el modelo Rasch, no solían aplicarse durante la construcción de las escalas de medición, por lo menos no el ámbito de la salud (Tennant et al. 2004, Brown 2006). Cabe aclarar, que el autor original de la escala no realizó ningún tipo de validación del instrumento. En el artículo publicado en 1994 se presentaron los motivos de la creación de la escala, el modo de selección de los ítems, los componentes y la derivación de las correspondientes puntuaciones, pero ningún dato de resultados medidos en pacientes (Richards et al. 1994). El primer estudio de validación de la parte auto-administrada de ASES (ASES-p) se publicó, años más tarde, en 2002. En este trabajo, la estimación de la consistencia interna de la escala, se basó en los 10 ítems de la función sin considerar el ítem del dolor (Michener et al. 2002). En los estudios de validación de esta escala a otros idiomas, se han aplicado análisis de componentes principales y factoriales exploratorios, de nuevo basados solo en los ítems de la función (Yahia et al. 2011, Piitulainen et al. 2014). No se ha conseguido encontrar ningún artículo que incorpore el ítem del dolor en estos análisis. Incorporación que sí se realizó en la validación realizada por nuestro grupo. Como parte de esta validación se aplicaron modelos de análisis factorial confirmatorio (AFC) y el modelo Rasch. Entre otras funciones, estos modelos permiten estudiar el número de factores latentes con los que se asocian los diferentes ítems de un cuestionario y establecer su dimensionalidad. A menudo estos modelos se aplican juntos ya que aportan visiones distintas sobre un mismo problema (Christensen et al. 2012). En este estudio se exploró la presencia de un único factor latente. Los resultados, presentados en el capítulo 6, apoyan la unidimensionalidad de ASES-p y, por lo tanto, la existencia de un sistema de puntuación común, basada en todos sus ítems (Vrotsou et al. 2016).

En este punto es pertinente realizar ciertas observaciones, tanto sobre la dimensionalidad como sobre el sistema de puntuación de la escala de interés. Aunque el ASES-p se presentó, y se está utilizando, como un cuestionario de dos dimensiones (dolor y función), el hecho que la dimensión del dolor se evalúa por un único ítem

implica, que por lo menos desde la perspectiva de los modelos utilizados, la presencia de dos factores latentes no se pudo comprobar, por lo tanto la bi-dimensionalidad de la escala tampoco se pudo confirmar con los métodos implementados (Brown 2006). Este hecho, que nos hace reflexionar y cuestionar la distribución de la puntuación total en dos sub-escalas. Asimismo, los datos de nuestro entorno no indican que el dolor sea el ítem más *difícil* para los participantes. Acciones como levantar 4.5kg por encima del hombro, o lanzar una pelota por encima de la cabeza presentaron más dificultad. Por último, los efectos estandarizados, relacionados con la sensibilidad al cambio, indicaron que la puntuación total y de la función, pueden distinguir entre sujetos mejorados y no mejorados con tamaños de efecto  $>1$ . Dicha distinción no fue tan clara para la sub-escala del dolor, cuyos respectivos efectos fueron de moderados a bajos. Aunque también cabe mencionar que las tres puntuaciones presentaron efectos más grandes en los pacientes que se sometieron a una cirugía de hombro, en comparaciones con pacientes sometidos a otros tratamientos. Asimismo la sub-escala del dolor presentó correlaciones más elevadas con el componente de dolor de las escalas CMS y SF-36, en comparación con otros.

Los datos de la *fiabilidad* test-retest, recopilados en un momento posterior mediante entrevistas telefónicas (capítulo 7), tampoco indicaron una buena reproducibilidad para la sub-escala de dolor. Debido a la ausencia de dolor en gran parte de la muestra, la concordancia exacta, entre las dos administraciones consecutivas, fue más bien elevada. A pesar de ello, los límites de concordancia al 95%, de esta sub-escala, fueron casi dos veces más anchos que los de la función, oscilando entre -13.2 hasta 11.9 puntos. Variación muy amplia para un rango de 0-50 puntos. Lo arriba mencionado pone en duda la distribución de la puntuación total de ASES-p en sus dos sub-escalas y el “peso” que el dolor tiene en esta distribución. Se subraya por lo tanto la necesidad de estudiar este aspecto con más detalle.

Nuestros resultados dirigen también la atención al último ítem de la escala, el ítem del deporte. A pesar de que ninguno de los análisis justificaría la exclusión de este ítem de la escala, se considera pertinente comentar sobre ciertos aspectos que pueden condicionar el buen funcionamiento del cuestionario en futuros estudios. En nuestro

entorno, los mayores de 65 años son menos probables en realizar actividades deportivas (Departamento de Salud del Gobierno Vasco 2007), hecho que implica que el ASES-p, en su formato actual, no se podría rellenar entero por un elevado porcentaje de esta población. Al no existir una imputación específica para el ítem en cuestión, o una variación de la escala para los sujetos que no responden al mismo, la no respuesta a este ítem puede implicar la exclusión de los análisis.

Sería interesante seguir avanzando en todas estas cuestiones, e intentar proporcionar soluciones que mejoren el funcionamiento de la escala. La puntuación total de ASES-p es una línea de investigación muy interesante. Los análisis factoriales y Rasch son herramientas que pueden permitir, aparte de lo hasta ahora realizado, la ponderación de los ítems y quizás indicar la derivación de una puntuación total más adecuada, o mejor ajustada, a los datos obtenidos de nuestro entorno. Otro método que se podría utilizar para realizar diferentes asignaciones de pesos, podría ser el método de correspondencias (*multiple correspondence analysis*). Este método, se puede usar para explorar la relación subyacente de las respuestas obtenidas. De modo resumido, se da el efecto Guttman que ordena las respuestas de los ítems y nos permite asignar pesos en función de este orden (Di Franco 2016). A pesar de que esta metodología todavía se encuentra en fase de desarrollo y, en principio, su naturaleza es exploratoria, podría ser interesante estudiar y comprobar la concordancia de los pesos asignados por los diferentes métodos arriba mencionados. Por otro lado, dada la falta de instrucciones sobre imputación de datos perdidos durante la administración de ASES-p, también sería relevante explorar diferentes modos de imputación de estos datos. Escalas como el SF-36 contemplan estos aspectos y cuentan con algoritmos de imputación (Saris-Baglana et al. 2011, Peyre et al. 2011). En relación al ítem del deporte, otra posibilidad interesante podría ser la creación de un ítem adaptable al nivel de actividad física que realizan los participantes. En este caso se estudiaría además la manera de homogeneizar las puntuaciones obtenidas, según dicho nivel, para que sean directamente comparables. Homogeneizaciones de este tipo se pueden estudiar por ejemplo, mediante el modelo Rasch aplicando un anclaje de puntuaciones (anchoring) (Bond & Fox 2007).

Aunque la escala ASES-p se desarrolló para ser aplicable a varias patologías del hombro, también sería de interés conocer si la escala es verdaderamente apta para todas ellas, y sobre todo para la patología de la inestabilidad. La publicación original de la escala completa, incluye una pregunta sobre inestabilidad. Aun así, las respuestas a esta pregunta no se acogen en la puntuación de ASES-p, tampoco se explica cómo se deberían implementar los datos en caso de una posible inestabilidad (Richards et al. 1994). Sería por lo tanto interesante comprobar este hecho. Existe la posibilidad de que el dolor agudo que se suele observar en esta patología, introduzca algún sesgo a las puntuaciones ASES-p, al igual que sucede con la escala CMS (Constant et al. 2008), aunque otros autores aseguran que la escala en cuestión es una candidata a considerar, en la evaluación de la inestabilidad de hombro (Plancher & Lipnick 2009).

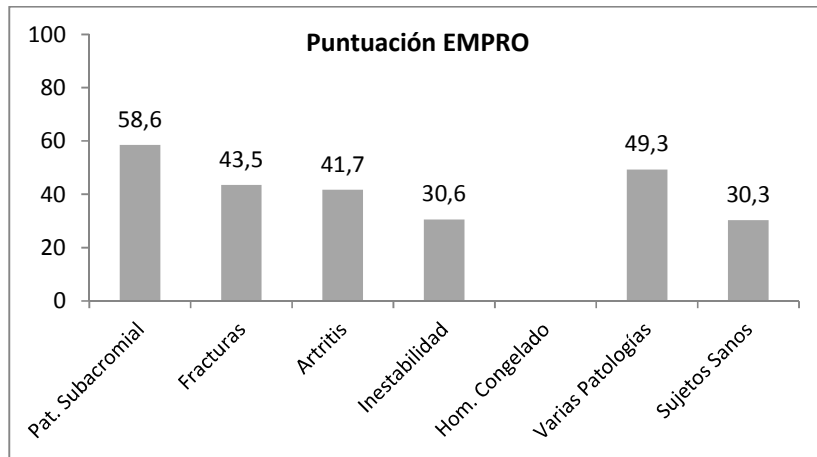
#### ***En relación al instrumento CMS***

Como se ya destaca en secciones anteriores del presente manuscrito, la escala Constant-Murley Score (CMS) es uno de los instrumentos más utilizados en la evaluación de hombro, a nivel mundial (Angst et al. 2011, Barra-Lopez 2007, Kirkley et al. 2003, Roy et al. 2010). Un dato indicativo son las más de 2000 citaciones de la publicación original, según el ISI Web of Knowledge (mayo 2017).

La escala CMS se ha tratado en dos partes de esta tesis. Primero como una escala administrada durante la recogida de datos para el proceso de la validación de ASES-p (capítulo 6) y en segundo lugar como objeto de una evaluación (capítulo 8). Los datos recopilados con el CMS, durante la validación de ASES-p, presentaron el comportamiento esperado. Las correlaciones entre las dos escalas y de sus componentes, cumplieron las hipótesis previamente establecidas, apoyando así la validez convergente. Eso, se podría quizás atribuir al que la mayoría de los participantes de nuestro estudio padecían una patología subacromial. A pesar del hecho de que la aplicación de CMS no supuso ningún problema especial en la realización de nuestro estudio, las referencias encontradas en la literatura mencionaban una serie de aspectos de esta escala, que nos llamaron la atención. Principalmente, nos extrañó que a pesar del hecho que el CMS se recomendaba por la *Sociedad Europea de Cirugía de Hombro y*

*Codo* (SECEC: <http://www.secec.org/>) y la *Sociedad Española de Cirugía de Hombro y Codo* (SECHC: <http://sehc.es/>), y a pesar de las numerosas publicaciones sobre el mismo, no se pudo encontrar ninguna validación adecuada de sus características. Aunque varios autores habían comentado sobre los inconvenientes de este instrumento, y otros habían desaconsejado su uso en ciertas patologías, tampoco se encontró una evaluación estandarizada, que permitiera establecer su funcionamiento en las diferentes patologías de la articulación, según la evidencia disponible en la literatura (Kirkley et al. 2003, Angst et al. 2011, Slobogean & Slobogean 2011). Una revisión sistemática del año 2010 estudió las características psicométricas de CMS, pero la evidencia que los autores encontraron en diferentes estudios, se expuso de modo narrativo (Roy et al. 2010). Cabe mencionar que durante el transcurso de nuestro proyecto otra publicación evaluó, mediante la herramienta COSMIN, varias escalas utilizadas en pacientes con patología subacromial, entre ellas el CMS (Huang et al. 2015). Los resultados de este trabajo se comentan en el capítulo actual, aunque el hecho que Huang et al. denomina el CMS “cuestionario auto-administrado”, genera ciertas dudas sobre la calidad de su trabajo.

El hecho de que el CMS no sea una escala auto-administrada, no significa que tenga necesariamente unas buenas propiedades métricas. A pesar de la creencia de ciertos investigadores, es pertinente que estas propiedades se evalúen (Booker et al. 2015, Barra-Lopez 2007). Por este motivo, nuestro grupo decidió valorar las propiedades de CMS, siguiendo otra vez una serie de pasos estandarizados, utilizando, de nuevo, la herramienta EMPRO. Los artículos finalmente incluidos en este proyecto, se agruparon en 5 grupos de patologías: *patología subacromial*, *fracturas*, *artritis*, *inestabilidad*, *hombro congelado*, y en 2 grupos adicionales: *varias patologías* y *sujetos sanos*. Las puntuaciones totales asignadas por EMPRO a cada grupo, fueron en general bajas y solo el grupo de la patología subacromial superó el mínimo valor aceptable por la herramienta (Figura 9.2). El grupo *varias patologías*, obtuvo la segunda mejor evaluación global, aunque con valor <50 puntos.



**Figura 9.2:** Puntuaciones EMPRO asignadas a la escala Constant-Murley Score, en siete grupos de patologías de hombro.

Cabe aquí mencionar, que la mayoría de los pacientes de este grupo padecían una patología subacromial. Por último, el grupo de los *sujetos sanos* recibió, en los atributos que se pudieron valorar, puntuaciones que sobrepasaban el punto de corte. Esto se debe a que la mayoría de los artículos en sujetos sanos estaban precisamente estudiando ciertas propiedades métricas de CMS. En relación a las diferentes propiedades métricas de la escala, las evaluaciones obtenidas indicaron lo siguiente. El *concepto y modelo de medida* no fue adecuado, sobre todo debido a la falta de justificación sobre la selección de los ítems, las dimensiones exploradas y la participación de la población diana en la creación de CMS. Autores anteriores han realizado observaciones parecidas (Conboy et al. 1996, Kirkley et al. 2003, Barra-Lopez 2007). La *consistencia interna* de la escala fue el atributo peor valorado, ya que información sobre el mismo se encontró solo en los grupos de patología subacromial y varias patologías. Eso llama la atención, teniendo sobre todo en cuenta que la estimación de la consistencia interna es uno de los datos más frecuentes en el ámbito de los cuestionarios de la salud, por lo menos para los cuestionarios auto-cumplimentados (Mc Dowell 2006). La falta de esta información podría tener una doble lectura. Por un lado, podría reflejar que este atributo no se perciba como una propiedad prioritaria para los clínicos. Pero también podría sugerir que ciertos autores evitan presentar este dato, por no haber obtenido una estimación deseable. Las pocas estimaciones de consistencia



interna que encontramos, entre los artículos revisados, presentaban valores alpha de Cronbach  $>0.60$ . La única excepción fue un artículo del grupo varias patologías, que presentó un alpha muy bajo, de 0.37 (Oh et al. 2009). Publicaciones anteriores han situado la consistencia interna entre 0.60-0.75 (Roy et al. 2010), mientras que otros concluye “*evidencia indeterminada*” sobre este atributo, en pacientes con patología subacromial (Huang et al. 2015).

En los artículos evaluados por nuestro estudio, no se consiguió información que apoyara la *reproducibilidad* de CMS en pacientes con artritis, inestabilidad, y hombro congelado. Según Roy et al. 2010, el CMS cuenta con excelente reproducibilidad en todos los estudios de su revisión. Pero esta conclusión se basa, sobre todo, en valores de correlación. Se conoce que el coeficiente de correlación no es un estadístico adecuado para valorar la reproducibilidad, ya que no capta las diferencias sistemáticas entre mediciones. Otros métodos estadísticos, como el coeficiente de correlación intraclase, o los límites de concordancia, son más recomendables para estudiar este atributo (Altman 1999). La *validez* de CMS recibió puntuaciones aceptables en los grupos de varias patologías y sujetos sanos. Pero solo el ítem EMPRO que evalúa la *validez convergente* y la *validez de grupos conocidos* (Hipótesis previas) recibió una puntuación elevada y constante en los grupos de patologías subacromial, fracturas, artritis, e inestabilidad. Otros autores han comentado que la *validez convergente* depende de las escalas utilizadas (Roy et al. 2010), y según otros, la información disponible sobre este atributo, en paciente con patología subacromial, no es concluyente (Huang et al. 2015).

La *sensibilidad al cambio* fue la propiedad mejor evaluada de CMS. Aunque esta característica no se pudo valorar para el hombro congelado y los sujetos sanos (en este caso por no ser relevante), los otros grupos recibieron entre 50 y 83.3 puntos EMPRO. Según Roy et al. 2010, el CMS tiene una excelente *sensibilidad al cambio*, salvo en la patología de la inestabilidad. Huang et al. 2015 concluye evidencia moderada sobre esta propiedad, en pacientes con patología subacromial. Nuestras evaluaciones, basadas en un mayor número de publicaciones que los trabajos arriba citados, sugieren que con excepción el grupo de la patología subacromial, más información hubiera sido deseable en el resto de patologías. La propiedad de la *interpretabilidad*, obtuvo evaluaciones

aceptables solo en los grupos subacromial y de varias patologías. Aun así, cabe destacar que esta propiedad fue, en todo caso, mejor valorada que la *consistencia interna*. Además, esta característica recibió puntuaciones más elevadas que la validez en los grupos de la patología subacromial, fracturas y artritis. Como ya se ha dicho anteriormente, la interpretabilidad es el grado en que la puntuación de un instrumento puede asignarse un significado fácilmente comprensible. Las valoraciones más elevadas de este atributo (aunque no necesariamente aceptables), en comparación con otras características, podrían reflejar que esta propiedad ha sido mejor estudiada que otras, en este ámbito. Este atributo no fue evaluado en los otros trabajos de revisión (Roy et al. 2010, Huang et al. 2015), aunque Roy et al. 2010, aconseja precaución con la interpretación de las puntuaciones de CMS.

Finalmente, los sub-atributos de la *facilidad de uso* presentaron resultados muy variables. Ninguno de ellos se pudo valorar para fracturas, hombro congelado y sujetos sanos. La *facilidad de uso para el entrevistado* no fue aceptable en artritis, inestabilidad, y varias patologías. Por otro lado la *carga de administración* no se consideró aceptable en los sujetos con inestabilidad. Otros autores han comentado sobre la dificultad de los pacientes con artritis en realizar la prueba de la fuerza de CMS (Christie et al. 2009). También han explicado que las particularidades de la inestabilidad de hombro, especialmente el dolor intermitente, hacen que el CMS no sea apto para esta condición (Conboy et al. 1996, Dawson et al. 1999, Constant et al. 2008, Plancher & Lipnick 2009).

Con relación a la prueba de la *fuerza* de CMS, es importante señalar la necesidad de aplicar un protocolo estandarizado para la valoración de la misma. La falta de instrucciones específicas sobre cómo realizar esta prueba, por parte de los autores originales, tuvo como resultado una enorme variación en el proceso de su realización. Este hecho, hace cuestionable la comparabilidad de los resultados obtenidos por CMS en diferentes proyectos. Además, las recomendaciones realizadas por el autor original en 2008, y los cambios propuestos en las respuestas de ciertos ítems (de categóricos a una escala visual analógica), aumentaron la existente variabilidad. En el año 2013 un grupo Danés, publicó un sencillo protocolo que podría constituir un primer paso, tanto para establecer una única versión de CMS, como para homogeneizar la realización del

componente de la fuerza (Ban et al. 2013). Dada la importancia que los aspectos arriba mencionados tienen en el resultado final de CMS, consideramos que la presentación de los datos de esta escala, debería siempre acompañarse con información precisa sobre su administración. La herramienta que se usa para la valoración de la fuerza, la posición del brazo, y el número de repeticiones son variables importantes que puede condicionar los resultados obtenidos (Moeller et al. 2014, Burrus et al. 2016, Constant et al. 2008). Pensamos que las revistas especializadas en este ámbito, y las sociedades correspondientes, también podrían ayudar hacia esta homogeneización, apoyando la difusión de un protocolo y una única versión de CMS.

Durante la realización de este proyecto se publicaron dos estudios de adaptación de la escala CMS a otros idiomas, uno a portugués brasileño y otra a turco. Ambos trabajos siguieron un proceso de adaptación cultural estandarizado y ambos adaptaron la versión modificada del instrumento CMS (Constant et al. 2008). Los autores del estudio brasileño realizaron además un modelo factorial exploratorio, con datos de n=110 pacientes con varias patologías de hombro. Según sus resultados la versión adaptada, cuanta con alta consistencia interna ( $\alpha=0,85$ ), mientras que todos sus ítems se asocian con un factor latente. En este trabajo no se estudió la reproducibilidad, ni la sensibilidad al cambio de la escala (Barreto et al. 2016). La adaptación turca, basada en datos de n=30 sujetos presento un CCI=0.87 y correlaciones  $\leq 0.48$  con las escalas ASES y SF-12 (Celik 2016). Cabe mencionar que aunque la SECHC proporciona la versión castellana de CMS, hasta la fecha (mayo 2017) no se ha podido encontrar ninguna validación de CMS al idioma diana.

Concluyendo, la evidencia disponible no apoyaría el uso de CMS en pacientes con fracturas, artritis, inestabilidad, y hombro congelado. La escala se puede usar en la valoración y el seguimiento de los pacientes con patología subacromial. Aunque incluso en este grupo, su consistencia interna y su validez no han sido adecuadamente estudiadas. Los resultados de nuestro proyecto, sugieren que el CMS no puede considerarse como regla de oro en la evaluación de hombro. Nuestros resultados podrían ser un punto de partida, para futuras líneas de investigación. Por un lado, la estructura subyacente de la escala, su dimensionalidad, y la utilidad de sus respectivos

Ítems son aspectos que requieren de estudio. En este campo, de utilidad podrían ser, de nuevo, los modelos de AFC y Rasch. Aparte de las propiedades del instrumento original, sería también interesante comprobar las propiedades métricas del CMS modificado, del ajustado por edad y sexo (Constant et al. 2008), o de la escala sin el componente de la fuerza (Othman & Taylor 2004). Consideramos que todavía hace falta mucha investigación en este campo.

## 10. Conclusiones generales

- Las herramientas de evaluación de las propiedades métricas, son de gran utilidad en el ámbito de las escalas de medición. Estandarizan la valoración de la evidencia y suponen un gran avance frente a la síntesis narrativa. Aumentan por lo tanto la calidad científica de los resultados obtenidos. Asimismo, consisten en una valiosa ayuda en el proceso de la recopilación de información sobre varias escalas, la comparación de la misma y la selección del instrumento más adecuado.
- La herramienta EMPRO, utilizada en el estudio de evaluación de diferentes escalas de CVRS, en pacientes con patologías de hombro, sugirió que las escalas ASES-p, SST y OSS, contaban, de manera global, con mejores propiedades psicométricas que el resto. La selección de ASES-p, para su posterior adaptación y validación al castellano se basó entre otros factores a su extensión de uso en el ámbito de la patología de hombro.
- La versión castellana de ASES-p es válida, fiable y sensible al cambio. A pesar de ello la evidencia, basada en los datos de nuestro entorno, no apoya la distribución de su puntuación total en dos sub-escalas. Concretamente la sensibilidad al cambio de la sub-

escala del dolor presenta estimaciones moderadas/bajas, mientras que los límites de concordancia de la misma no se pueden considerar aceptables.

- Tanto los análisis factoriales confirmatorios, como el modelo Rasch, apoyan la unidimensionalidad de ASES-p. Aun así el ítem del deporte se tendría que considerar con más detalle. Este ítem presentó correlaciones  $<0.30$  con otros dos ítems de la escala, su ajuste al modelo Rasch no fue óptimo y, además, fue el ítem con la mayor cantidad de datos perdidos, sobre todo en sujetos mayores de 65 años.
- La existencia o no de dos sub-escalas es un aspecto sobre el cual tendremos que profundizar. Asimismo, una posible ponderación de los ítems de ASES-p, mediante la asignación de pesos, es probable que mejore el funcionamiento del instrumento. También el modo de imputación de los valores perdidos, o incluso la creación de una variación de la escala para ser aplicable a sujetos con diferentes niveles de actividad física, pueden constituir líneas para futuros proyectos de investigación.
- Aunque herramientas, como EMPRO, se desarrollaron para valorar las escalas auto-complimentadas por los pacientes, nuestra experiencia avala su uso con medidas de otro tipo. En esta tesis EMPRO se utilizó para evaluar las propiedades métricas de escala CMS, que combina datos subjetivos con datos de observación clínica.
- La escala CMS, en contra de una parte de la opinión clínica, debería ser evaluada de manera rigurosa, igual que los instrumentos auto-administrados. No existe ningún motivo que impida este tipo de evaluación, sobre todo cuando se trata de una escala que muchos perciben como regla de oro en la evaluación de la patología de hombro.
- Los resultados de nuestro proyecto no apoyan el uso de CMS en la evaluación de todas las patologías de la articulación de hombro. La evidencia disponible sugiere la escala tiene un buen funcionamiento en pacientes con patología subacromial. Aun así, la consistencia interna y la validez de la misma son atributos que requieren estudio adicional. Finalmente, cabe mencionar que hacen falta protocolos estandarizados para homogeneizar la aplicación de CMS y asegurar la comparabilidad de las puntuaciones obtenidas en diferentes estudios.

## Bibliografía

1. Aaronson N, Alonso J, Burnam A, Lohr KN, Patrick DL, Perrin E et al.: Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res* 2002, 11: 193-205.
2. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ et al.: The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst* 1993, 85: 365-376.
3. Abreo AP, Glidden D, Painter P, Lea J, Herzog CA, Kutner NG et al.: Association of physical function with predialysis blood pressure in patients on hemodialysis. *BMC Nephrol* 2014, 15: 177.
4. Alonso J: [The measurement of health related-quality of life in clinical research and practice]. *Gac Sanit* 2000, 14: 163-167.
5. Alonso J, Regidor E, Barrio G, Prieto L, Rodríguez C, de la Fuente L: [Population reference values of the Spanish version of the Health Questionnaire SF-36]. *Med Clin (Barc)* 1998, 111: 410-416.
6. Altman DG: *Practical statistics for medical research*. Boca Raton, Florida: Chapman & Hall; 1999.
7. 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: 185-187.
8. American Academy of Orthopaedic Surgeons. *Shoulder Joint Replacement*. 2013. Ref Type: Online Source
9. Anastasi A: Evolving concepts of test validation. *Ann Rev Psychol* 1986, 37: 1-15.
10. Andrade Ortega JA, gado Martínez AD, Almecija RR: [Validation of a Spanish version of the Neck Disability Index]. *Med Clin (Barc)* 2008, 130: 85-89.
11. Andresen EM: Criteria for assessing the tools of disability outcomes research. *Arch Phys Med Rehabil* 2000, 81: S15-S20.
12. Angst F, Goldhahn J, Drerup S, Aeschlimann A, Schwyzer HK, Simmen BR: Responsiveness of six outcome assessment instruments in total shoulder arthroplasty. *Arthritis Rheum* 2008, 59: 391-398.
13. Angst F, Pap G, Mannion AF, Herren DB, Aeschlimann A, Schwyzer HK et al.: Comprehensive assessment of clinical outcome and quality of life after total shoulder arthroplasty: usefulness and validity of subjective outcome measures. *Arthritis Rheum* 2004, 51: 819-828.

14. Angst F, Schwyzer HK, Aeschlimann A, Simmen BR, 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 (Hoboken)* 2011, 63 Suppl 11: S174-S188.
15. Arroyo-Aljaro R, Gonzalez-Viejo MA: Validación al castellano del Wheelchair Users Shoulder Pain Index (WUSPI). *Rehabilitacion (Madr)* 2009, 43: 2-9.
16. Arthritis Research UK. Shoulder and elbow replacement. 2016. Ref Type: Online Source
17. Badia X, Baro E: Cuestionarios de salud en España y su uso en atención primaria. *Atención Primaria* 2001, 28: 349-356.
18. Baker P, Nanda R, Goodchild L, Finn P, Rangan A: A comparison of the Constant and Oxford shoulder scores in patients with conservatively treated proximal humeral fractures. *J Shoulder Elbow Surg* 2008, 17: 37-41.
19. Baldwin AL, Fullmer K, Schwartz GE: Comparison of physical therapy with energy healing for improving range of motion in subjects with restricted shoulder mobility. *Evid Based Complement Alternat Med* 2013, 2013: 329731.
20. Ban I, Troelsen A, Christiansen DH, Svendsen SW, Kristensen MT: Standardised test protocol (Constant Score) for evaluation of functionality in patients with shoulder disorders. *Dan Med J* 2013, 60: A4608.
21. Bankes MJ, Crossman JE, Emery RJ: A standard method of shoulder strength measurement for the Constant score with a spring balance. *J Shoulder Elbow Surg* 1998, 7: 116-121.
22. Barofsky I: Can quality or quality-of-life be defined? *Qual Life Res* 2012, 21: 625-631.
23. Barra-Lopez ME: El test de Constant-Murley. Una revision de sus características. *Rehabilitacion (Madr)* 2007, 41: 228-235.
24. Barreto RP, Barbosa ML, Balbinotti MA, Mothes FC, da Rosa LH, Silva MF: The Brazilian version of the Constant-Murley Score (CMS-BR): convergent and construct validity, internal consistency, and unidimensionality. *Rev Bras Ortop* 2016, 51: 515-520.
25. Baumgarten KM, Helsen E: Does chondrolysis occur after corticosteroid-analgesic injections? An analysis of patients treated for adhesive capsulitis of the shoulder. *J Shoulder Elbow Surg* 2016, 25: 890-897.
26. Baumgarten KM, Helsen E: Does chondrolysis occur after corticosteroid-analgesic injections? An analysis of patients treated for adhesive capsulitis of the shoulder. *J Shoulder Elbow Surg* 2016, 25: 890-897.
27. Beaton DE, Richards RR: Measuring function of the shoulder. A cross-sectional comparison of five questionnaires. *J Bone Joint Surg Am* 1996, 78: 882-890.



28. Beaton DE, Wright JG, Katz JN: Development of the QuickDASH: comparison of three item-reduction approaches. *J Bone Joint Surg Am* 2005, 87: 1038-1046.
29. Black N: Patient reported outcome measures could help transform healthcare. *BMJ* 2013, 346: f167.
30. Bland JM. How can I decide the sample size for a repeatability study. 2010.
31. Bland JM, Altman DG: Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986, 1: 307-310.
32. Bland JM, Altman DG: Cronbach's alpha. *BMJ* 1997, 314: 572.
33. Blonna D, Scelsi M, Marini E, Bellato E, Tellini A, Rossi R et al.: Can we improve the reliability of the Constant-Murley score? *J Shoulder Elbow Surg* 2012, 21: 4-12.
34. Bohannon RW: Hand-grip dynamometry predicts future outcomes in aging adults. *J Geriatr Phys Ther* 2008, 31: 3-10.
35. Bombardier C, Tugwell P: Methodological considerations in functional assessment. *J Rheumatol Suppl* 1987, 14 Suppl 15: 6-10.
36. Bond TG, Fox CM: *Applying The Rasch Model: Fundamental Measurement in the Human Sciences*, Second edition edn. 2007.
37. Bongers PM: The cost of shoulder pain at work. *BMJ* 2001, 322: 64-65.
38. Bonner NS, O'Halloran PD, Bernhardt J, Cumming TB: Developing the Stroke Exercise Preference Inventory (SEPI). *PLoS One* 2016, 11: e0164120.
39. Booker S, Alfahad N, Scott M, Gooding B, Wallace WA: Use of scoring systems for assessing and reporting the outcome results from shoulder surgery and arthroplasty. *World J Orthop* 2015, 6: 244-251.
40. Bot SD, Terwee CB, Van Der Windt DA, Bouter LM, Dekker J, de Vet HC: Clinimetric evaluation of shoulder disability questionnaires: a systematic review of the literature. *Ann Rheum Dis* 2004, 63: 335-341.
41. Brazier JE, Harper R, Jones NM, O'Cathain A, Thomas KJ, Usherwood T et al.: Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *BMJ* 1992, 305: 160-164.
42. Breslow L: Health measurement in the third era of health. *Am J Public Health* 2006, 96: 17-19.
43. Brown TA: *Confirmatory Factor Analysis for Applied Research*. The Guilford Press; 2006.
44. Brown TA, White KS, Barlow DH: A psychometric reanalysis of the Albany Panic and Phobia Questionnaire. *Behav Res Ther* 2005, 43: 337-355.
45. Bullinger M, Alonso J, Apolone G, Leplege A, Sullivan M, Wood-Dauphinee S et al.: Translating health status questionnaires and evaluating their quality: the IQOLA Project approach. *International Quality of Life Assessment. J Clin Epidemiol* 1998, 51: 913-923.

46. Burrus C, Deriaz O, Luthi F, Konzelmann M: Role of pain in measuring shoulder strength abduction and flexion with the Constant-Murley score. *Ann Phys Rehabil Med* 2016.
47. Castellote Olivito JM, Huertas Olmedo F, Sanegre Llopis MT, Amate Blanco JM, Conde Olasagasti JL. Prótesis de hombro en Indicaciones de Procesos degenerativos o traumatológicos. Valoración de calidad de vida y actividad funcional como determinantes de la efectividad de esta sustitución articular. 25. 2000. Informe Técnico, Agencia de Evaluación de Tecnologías Sanitarias (AETS) del Instituto de Salud Carlos III, Ministerio de Sanidad y Consumo.
48. Celik D: Turkish version of the modified Constant-Murley score and standardized test protocol: reliability and validity. *Acta Orthop Traumatol Turc* 2016, 50: 69-75.
49. Celik D, Atalar AC, Demirhan M, Dirican A: Translation, cultural adaptation, validity and reliability of the Turkish ASES questionnaire. *Knee Surg Sports Traumatol Arthrosc* 2013, 21: 2184-2189.
50. Cheong SK, Johnston LM: Systematic review of self-concept measures for primary school aged children with cerebral palsy. *Res Dev Disabil* 2013, 34: 3566-3575.
51. Christensen KB, Engelhard G, Salzberger T. Ask the experts: Rasch vs. Factor Analysis. *Rasch Measurement Transactions* 26[3], 1373-1378. 2012.
52. Christie A, Hagen KB, Mowinckel P, Dagfinrud H: Methodological properties of six shoulder disability measures in patients with rheumatic diseases referred for shoulder surgery. *J Shoulder Elbow Surg* 2009, 18: 89-95.
53. Cid-Ruzafa J, Damian-Moreno J: [Disability evaluation: Barthel's index]. *Rev Esp Salud Publica* 1997, 71: 127-137.
54. Clifford PE, Mallon WJ: Sports after total joint replacement. *Clin Sports Med* 2005, 24: 175-186.
55. Coghlan JA, Bell SN, Forbes A, Buchbinder R: Comparison of self-administered University of California, Los Angeles, shoulder score with traditional University of California, Los Angeles, shoulder score completed by clinicians in assessing the outcome of rotator cuff surgery. *J Shoulder Elbow Surg* 2008, 17: 564-569.
56. Cohen J: A power primer. *Psychol Bull* 1992, 112: 155-159.
57. Conboy VB, Morris RW, Kiss J, Carr AJ: An evaluation of the Constant-Murley shoulder assessment. *J Bone Joint Surg Br* 1996, 78: 229-232.
58. Constant CR: Age related recovery of shoulder function after injury [MCh thesis<sup>1</sup>]. Cork, Ireland: University College; 1986.
59. Constant CR, Gerber C, Emery RJ, Sojbjerg JO, Gohlke F, Boileau P: A review of the Constant score: modifications and guidelines for its use. *J Shoulder Elbow Surg* 2008, 17: 355-361.
60. Constant CR, Murley AH: A clinical method of functional assessment of the shoulder. *Clinical Orthopaedics & Related Research* 1987, 160-164.

61. Cook KF, Gartsman GM, Roddey TS, Olson SL: The measurement level and trait-specific reliability of 4 scales of shoulder functioning: an empiric investigation. *Arch Phys Med Rehabil* 2001, 82: 1558-1565.
62. Cook KF, Roddey TS, Gartsman GM, Olson SL: Development and psychometric evaluation of the Flexilevel Scale of Shoulder Function. *Med Care* 2003, 41: 823-835.
63. Coons SJ, Rao S, Keininger DL, Hays RD: A comparative review of generic quality-of-life instruments. *Pharmacoeconomics* 2000, 17: 13-35.
64. Corniola MV, Stienen MN, Joswig H, Smoll NR, Schaller K, Hildebrandt G et al.: Correlation of pain, functional impairment, and health-related quality of life with radiological grading scales of lumbar degenerative disc disease. *Acta Neurochir (Wien)* 2016, 158: 499-505.
65. Croft P, Pope D, Zonca M, O'Neill T, Silman A: Measurement of shoulder related disability: results of a validation study. *Ann Rheum Dis* 1994, 53: 525-528.
66. CRONBACH LJ, MEEHL PE: Construct validity in psychological tests. *Psychol Bull* 1955, 52: 281-302.
67. Curtis KA, Roach KE, Applegate EB, Amar T, Benbow CS, Genecco TD et al.: Reliability and validity of the Wheelchair User's Shoulder Pain Index (WUSPI). *Paraplegia* 1995, 33: 595-601.
68. Cvetanovich GL, Chalmers PN, Streit JJ, Romeo AA, Nicholson GP: Patients Undergoing Total Shoulder Arthroplasty on the Dominant Extremity Attain Greater Postoperative ROM. *Clin Orthop Relat Res* 2015, 473: 3221-3225.
69. Dawson J, Fitzpatrick R, Carr A: Questionnaire on the perceptions of patients about shoulder surgery. *J Bone Joint Surg Br* 1996, 78: 593-600.
70. Dawson J, Fitzpatrick R, Carr A: The assessment of shoulder instability. The development and validation of a questionnaire. *J Bone Joint Surg Br* 1999, 81: 420-426.
71. Departamento de Salud del Gobierno Vasco. Encuesta de Salud del País Vasco. [http://www.osakidetza.euskadi.eus/r85-ckpubl01/es/contenidos/informacion/encuesta\\_salud/es\\_escav/encuesta\\_salud.html](http://www.osakidetza.euskadi.eus/r85-ckpubl01/es/contenidos/informacion/encuesta_salud/es_escav/encuesta_salud.html) . 2007. 22-4-2016. Ref Type: Electronic Citation
72. Dere D, Paker N, Soy BD, Tekdos DD: Effect of body mass index on functional recovery after total knee arthroplasty in ambulatory overweight or obese women with osteoarthritis. *Acta Orthop Traumatol Turc* 2014, 48: 117-121.
73. DeVon HA, Block ME, Moyle-Wright P, Ernst DM, Hayden SJ, Lazzara DJ et al.: A psychometric toolbox for testing validity and reliability. *J Nurs Scholarsh* 2007, 39: 155-164.
74. Dewitte K, Fierens C, Stockl D, Thienpont LM: Application of the Bland-Altman plot for interpretation of method-comparison studies: a critical investigation of its practice. *Clin Chem* 2002, 48: 799-801.

75. Di Franco G: Multiple correspondence analysis: one only or several techniques? *Qual Quant* 2016, 50: 1299-1315.
76. Escamilla RF, Yamashiro K, Paulos L, Andrews JR: Shoulder muscle activity and function in common shoulder rehabilitation exercises. *Sports Med* 2009, 39: 663-685.
77. European Society for Surgery of the Shoulder and the Elbow. Constant Score. 2016. 8-4-2016. Ref Type: Online Source
78. Ferguson CJ: An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice* 2009, 40: 532.
79. Flora DB, Labrish C, Chalmers RP: Old and new ideas for data screening and assumption testing for exploratory and confirmatory factor analysis. *Front Psychol* 2012, 3: 55.
80. Fong SS, Ng SS, Luk WS, Chung JW, Chung LM, Tsang WW et al.: Shoulder Mobility, Muscular Strength, and Quality of Life in Breast Cancer Survivors with and without Tai Chi Qigong Training. *Evid Based Complement Alternat Med* 2013, 2013: 787169.
81. Forero CG, Maydeu-Olivares A, Gallardo-Pujol D: Factor analysis with ordinal indicators: a Monte Carlo study comparing DWLS and ULS estimation. *Structural Equation Modelling* 2009, 16: 625-641.
82. Garin O, Herdman M, Vilagut G, Ferrer M, Ribera A, Rajmil L et al.: Assessing health-related quality of life in patients with heart failure: a systematic, standardized comparison of available measures. *Heart Fail Rev* 2013.
83. Garratt A, Schmidt L, Mackintosh A, Fitzpatrick R: Quality of life measurement: bibliographic study of patient assessed health outcome measures. *BMJ* 2002, 324: 1417.
84. Gellman M, Turner J. *Encyclopedia of Behavioural Medicine*. 2013. Springer New York.
85. Giesinger JM, Kesterke N, Hamilton DF, Holzner B, Jost B, Giesinger K: Development of an item list to assess the forgotten joint concept in shoulder patients. *BMC Musculoskelet Disord* 2015, 16: 67.
86. Gill TM: Assessment of function and disability in longitudinal studies. *J Am Geriatr Soc* 2010, 58 Suppl 2: S308-S312.
87. Ginés A, Torrens C, Solano A, uiz A, artínez A, aceres E: Influencia de la versión protésica y de la consolidación del troquíter en el resultado funcional de las hemiartroplastias implantadas por fractura de húmero proximal. *Revista de Ortopedia y Traumatología* 2007, 51: 7-14.
88. Godfrey J, Hamman R, Lowenstein S, Briggs K, Kocher M: Reliability, validity, and responsiveness of the simple shoulder test: psychometric properties by age and injury type. *J Shoulder Elbow Surg* 2007, 16: 260-267.

89. Goldhahn J, Angst F, Drerup S, Pap G, Simmen BR, Mannion AF: Lessons learned during the cross-cultural adaptation of the American Shoulder and Elbow Surgeons shoulder form into German. *J Shoulder Elbow Surg* 2008, 17: 248-254.
90. Guerra de Hoyos JA, Andres Martin MC, Baena de LE, Vigarra LM, Molina LT, Verdugo Morilla FA et al.: Randomised trial of long term effect of acupuncture for shoulder pain. *Pain* 2004, 112: 289-298.
91. Guion RM: On trinitarian doctrines of validity. *Professional Psychology* 1980, 11: 385-389.
92. Guralnik JM, Ferrucci L, Pieper CF, Leveille SG, Markides KS, Ostir GV et al.: Lower extremity function and subsequent disability: consistency across studies, predictive models, and value of gait speed alone compared with the short physical performance battery. *J Gerontol A Biol Sci Med Sci* 2000, 55: M221-M231.
93. Guyatt GH: Making sense of quality of life data. *Med Care* 2000, 38(suppl):II: 175-179.
94. Guyatt GH, Feeny DH, Patrick DL: Measuring health-related quality of life. *Ann Intern Med* 1993, 118: 622-629.
95. Hazuda HP, Dhanda R, Owen SV, Lichtenstein MJ: Development and validation of a performance-based measure of upper extremity functional limitation. *Aging Clin Exp Res* 2005, 17: 394-401.
96. Hegedus EJ, Goode A, Campbell S, Morin A, Tamaddoni M, Moorman CT, III et al.: Physical examination tests of the shoulder: a systematic review with meta-analysis of individual tests. *Br J Sports Med* 2008, 42: 80-92.
97. Herdman M, Badia X, Berra S: [EuroQol-5D: a simple alternative for measuring health-related quality of life in primary care]. *Aten Primaria* 2001, 28: 425-430.
98. Hervas MT, Navarro Collado MJ, Peiro S, Rodrigo Perez JL, Lopez MP, Martinez T, I: [Spanish version of the DASH questionnaire. Cross-cultural adaptation, reliability, validity and responsiveness]. *Med Clin (Barc )* 2006, 127: 441-447.
99. Hilditch JR, Lewis J, Peter A, van MB, Ross A, Franssen E et al.: A menopause-specific quality of life questionnaire: development and psychometric properties. *Maturitas* 1996, 24: 161-175.
100. Huang H, Grant JA, Miller BS, Mirza FM, 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.
101. Hudak PL, Amadio PC, Bombardier C: Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand) [corrected]. The Upper Extremity Collaborative Group (UECG). *Am J Ind Med* 1996, 29: 602-608.
102. Husted JA, Cook RJ, Farewell VT, Gladman DD: Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol* 2000, 53: 459-468.

103. Karlsson J, Taft C, Sjostrom L, Torgerson JS, Sullivan M: Psychosocial functioning in the obese before and after weight reduction: construct validity and responsiveness of the Obesity-related Problems scale. *Int J Obes Relat Metab Disord* 2003, 27: 617-630.
104. Kirkley A, Alvarez C, Griffin S: The development and evaluation of a disease-specific quality-of-life questionnaire for disorders of the rotator cuff: The Western Ontario Rotator Cuff Index. *Clin J Sport Med* 2003, 13: 84-92.
105. Kirkley A, Griffin S, Dainty K: Scoring systems for the functional assessment of the shoulder. *Arthroscopy* 2003, 19: 1109-1120.
106. Kirkley A, Griffin S, McLintock H, Ng L: The development and evaluation of a disease-specific quality of life measurement tool for shoulder instability. The Western Ontario Shoulder Instability Index (WOSI). *Am J Sports Med* 1998, 26: 764-772.
107. Knaut LA, Moser AD, Melo SA, Richards RR: Translation and cultural adaptation to the portuguese language of the American Shoulder and Elbow Surgeons Standardized Shoulder assessment form (ASES) for evaluation of shoulder function. *Rev Bras Reumatol* 2010, 50: 176-189.
108. Kocher MS, Horan MP, Briggs KK, Richardson TR, O'Holleran J, Hawkins RJ: Reliability, validity, and responsiveness of the American Shoulder and Elbow Surgeons subjective shoulder scale in patients with shoulder instability, rotator cuff disease, and glenohumeral arthritis. *J Bone Joint Surg Am* 2005, 87: 2006-2011.
109. Kohn D, Geyer M: The subjective shoulder rating system. *Arch Orthop Trauma Surg* 1997, 116: 324-328.
110. Kotecha D, Chudasama R, Lane DA, Kirchof P, Lip GY: Atrial fibrillation and heart failure due to reduced versus preserved ejection fraction: A systematic review and meta-analysis of death and adverse outcomes. *Int J Cardiol* 2016, 203: 660-666.
111. Kovacs FM, Llobera J, Gil Del Real MT, Abreira V, Gestoso M, Fernandez C et al.: Validation of the spanish version of the Roland-Morris questionnaire. *Spine (Phila Pa 1976 )* 2002, 27: 538-542.
112. Kuijpers T, van der Windt DA, van der Heijden GJ, Twisk JW, Vergouwe Y, Bouter LM: A prediction rule for shoulder pain related sick leave: a prospective cohort study. *BMC Musculoskeletal Disorders* 2006, 7.
113. L'Insalata JC, Warren RF, Cohen SB, Altchek DW, Peterson MG: A self-administered questionnaire for assessment of symptoms and function of the shoulder. *J Bone Joint Surg Am* 1997, 79: 738-748.
114. Largacha M, Parsons IM, Campbell B, Titelman RM, Smith KL, Matsen F, III: Deficits in shoulder function and general health associated with sixteen common shoulder diagnoses: a study of 2674 patients. *J Shoulder Elbow Surg* 2006, 15: 30-39.

115. Lee J, Lee EH, Moon SH: A systematic review of measurement properties of the instruments measuring health-related quality of life in patients with irritable bowel syndrome. *Qual Life Res* 2016, 25: 2985-2995.
116. Leggin BG, Lannotti J: Shoulder outcome measurement. In *Disorders of the Shoulder: Diagnosis and Management*. Edited by Lannotti J, Williams G. Philadelphia: PA: Lippincott, Williams & Wilkins; 1999:1024-1040.
117. Leggin BG, Michener LA, Shaffer MA, Brenneman SK, Iannotti JP, Williams GR, Jr.: The Penn shoulder score: reliability and validity. *J Orthop Sports Phys Ther* 2006, 36: 138-151.
118. Leggin BG, Shaffer MA, Neuman RM, Williams GR, Jr.: Relationship of the Penn Shoulder Score with measures of range of motion and strength in patients with shoulder disorders: A preliminary report. *The University of Pennsylvania Orthopaedic Journal* 2003, 16: 39-44.
119. Levy O, Haddo O, Massoud S, Mullett H, Atoun E: A patient-derived Constant-Murley score is comparable to a clinician-derived score. *Clin Orthop Relat Res* 2014, 472: 294-303.
120. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP et al.: The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol* 2009, 62: e1-34.
121. Lillkrona U: How should we use the Constant Score?--A commentary. *J Shoulder Elbow Surg* 2008, 17: 362-363.
122. Linacre JM. *A user's guide to Winsteps / Ministep Rasch-Model Computer Programs*. 2011.
123. Lippitt S, Harryman D, Matsen F: A practical tool for evaluating function: The simple shoulder test. In *The Shoulder: A balance of mobility and stability*. Edited by Matsen F, F F, Hawkins R. Rosemont, IL: The American Academy of Orthopaedic Surgeons; 1993:501-518.
124. Liu AH, Zeiger R, Sorkness C, Mahr T, Ostrom N, Burgess S et al.: Development and cross-sectional validation of the Childhood Asthma Control Test. *J Allergy Clin Immunol* 2007, 119: 817-825.
125. Lo IK, Griffin S, Kirkley A: The development of a disease-specific quality of life measurement tool for osteoarthritis of the shoulder: The Western Ontario Osteoarthritis of the Shoulder (WOOS) index. *Osteoarthritis Cartilage* 2001, 9: 771-778.
126. Lohr KN, Aaronson NK, Alonso J, Burnam MA, Patrick DL, Perrin EB et al.: Evaluating quality-of-life and health status instruments: development of scientific review criteria. *Clin Ther* 1996, 18: 979-992.
127. Longo UG, Vasta S, Maffulli N, Denaro V: Scoring systems for the functional assessment of patients with rotator cuff pathology. *Sports Med Arthrosc* 2011, 19: 310-320.

128. Lorenz C, Cheville AL, Amer H, Kotajarvi BR, Stegall MD, Petterson TM et al.: Relationship between pre-transplant physical function and outcomes after kidney transplant Elizabeth. *Clin Transplant* 2017.
129. Luime JJ, Koes BW, Hendriksen IJ, Burdorf A, Verhagen AP, Miedema HS et al.: Prevalence and incidence of shoulder pain in the general population; a systematic review. *Scand J Rheumatol* 2004, 33: 73-81.
130. Luque-Suarez A, Rondon-Ramos A, Fernandez-Sanchez M, Roach KE, Morales-Asencio JM: Spanish version of SPADI (shoulder pain and disability index) in musculoskeletal shoulder pain: a new 10-items version after confirmatory factor analysis. *Health Qual Life Outcomes* 2016, 14: 32.
131. Lusardi MM, Fritz S, Middleton A, Allison L, Wingood M, Phillips E et al.: Determining Risk of Falls in Community Dwelling Older Adults: A Systematic Review and Meta-analysis Using Posttest Probability. *J Geriatr Phys Ther* 2017, 40: 1-36.
132. Mahoney FI, Barthel DW: Functional Evaluation: The Barthel Index. *Md State Med J* 1965, 14: 61-65.
133. Mahoney FI, Wood OH, Barthel DW: Rehabilitation of chronically ill patients: the influence of complications on the final goal. *South Med J* 1958, 51: 605-609.
134. Mangen MJ, Huijts SM, Bonten MJ, de Wit GA: The impact of community-acquired pneumonia on the health-related quality-of-life in elderly. *BMC Infect Dis* 2017, 17: 208.
135. Maratia S, Cedillo S, Rejas J: Assessing health-related quality of life in patients with breast cancer: a systematic and standardized comparison of available instruments using the EMPRO tool. *Qual Life Res* 2016, 25: 2467-2480.
136. Marin-Gomez M, Navarro-Collado MJ, Peiro S, Trenor-Gomis C, Paya-Rubio A, Bernal-Delgado E et al.: [The quality of care in shoulder pain. A medical audit]. *Gac Sanit* 2006, 20: 116-123.
137. Mc Dowell I: *Measuring Health: A Guide to Rating Scales and Questionnaires*, Third Edition edn. Oxford University Press; 2006.
138. McDowell I, Jenkinson C: Development standards for health measures. *J Health Serv Res Policy* 1996, 1: 238-246.
139. Medina-Porqueres I, Cantero-Tellez R, Gomez-Calvente M, Matsen F (Eds): Prospective, Multicenter Study Protocol to Adaptate and Validate the Simple Shoulder Test to Spanish Population. In *Journal of Novel Physiotherapies* 2015, 5: 275.
140. Membrilla-Mesa MD, Tejero-Fernandez 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: 411-416.
141. Messick S: Standards of validity and the validity of standards in performance assessment. *Educational measurement: Issues and practice* 1995, 14: 5-8.



142. 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: 68-76.
143. Michener LA, McClure PW, Sennett BJ: American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form, patient self-report section: reliability, validity, and responsiveness. *J Shoulder Elbow Surg* 2002, 11: 587-594.
144. Middel B, van SE: Statistical significant change versus relevant or important change in (quasi) experimental design: some conceptual and methodological problems in estimating magnitude of intervention-related change in health services research. *Int J Integr Care* 2002, 2: e15.
145. Mitchell LE, Ziviani J, Oftedal S, Boyd RN: A systematic review of the clinimetric properties of measures of habitual physical activity in primary school aged children with cerebral palsy. *Res Dev Disabil* 2013, 34: 2419-2432.
146. Moeller AD, Thorsen RR, Torabi TP, Bjoerkman AS, Christensen EH, Maribo T et al.: The Danish version of the modified Constant-Murley shoulder score: reliability, agreement, and construct validity. *J Orthop Sports Phys Ther* 2014, 44: 336-340.
147. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL et al.: The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010, 19: 539-549.
148. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL et al.: The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010, 63: 737-745.
149. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL. The COSMIN checklist manual. [www.cosmin.nl](http://www.cosmin.nl) . 2012. 15-12-2011.
150. Newman AB, Cauley JA: The epidemiology of aging. Edited by Springer. 2012.
151. Oh JH, Jo KH, Kim WS, Gong HS, Han SG, Kim YH: Comparative evaluation of the measurement properties of various shoulder outcome instruments. *Am J Sports Med* 2009, 37: 1161-1168.
152. Osakidetza Servicio Vasco de Salud DdAS. Proyecto estrategico 8: Competencias avanzadas de enfermería. 2017. 18-3-2017.
153. Othman A, Taylor G: Is the constant score reliable in assessing patients with frozen shoulder? 60 shoulders scored 3 years after manipulation under anaesthesia. *Acta Orthop Scand* 2004, 75: 114-116.
154. Paananen M, Taimela S, Auvinen J, Tammelin T, Zitting P, Karppinen J: Impact of self-reported musculoskeletal pain on health-related quality of life among young adults. *Pain Med* 2011, 12: 9-17.
155. Padua R, Padua L, Ceccarelli E, Bondi R, Alviti F, Castagna A: Italian version of ASES questionnaire for shoulder assessment: cross-cultural adaptation and validation. *Musculoskelet Surg* 2010, 94 Suppl 1: S85-S90.

156. Pallant JF, Tennant A: An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). *Br J Clin Psychol* 2007, 46: 1-18.
157. Patte D.. The first open congress of the European Society of Surgery of the Shoulder and Elbow [SECEC]. Directions for the use of the index severity for painful and/or chronically disabled shoulders. 36-41. 1987.
158. Paul A, Lewis M, Shadforth MF, Croft PR, Van Der Windt DA, Hay EM: A comparison of four shoulder-specific questionnaires in primary care. *Ann Rheum Dis* 2004, 63: 1293-1299.
159. Peyre H, Lepage A, Coste J: Missing data methods for dealing with missing items in quality of life questionnaires. A comparison by simulation of personal mean score, full information maximum likelihood, multiple imputation, and hot deck techniques applied to the SF-36 in the French 2003 decennial health survey. *Qual Life Res* 2011, 20: 287-300.
160. Piitulainen K, Paloneva J, Ylinen J, Kautiainen H, Hakkinen A: Reliability and validity of the Finnish version of the American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form, patient self-report section. *BMC Musculoskelet Disord* 2014, 15: 272.
161. Placzek JD, Lukens SC, Badalanmenti S, Roubal PJ, Freeman DC, Walleman KM et al.: Shoulder outcome measures: a comparison of 6 functional tests. *Am J Sports Med* 2004, 32: 1270-1277.
162. Plancher KD, Lipnick SL: Analysis of evidence-based medicine for shoulder instability. *Arthroscopy* 2009, 25: 897-908.
163. Podsiadlo D, Richardson S: The timed "Up & Go": a test of basic functional mobility for frail elderly persons. *J Am Geriatr Soc* 1991, 39: 142-148.
164. Preston CC, Colman AM: Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychol (Amst)* 2000, 104: 1-15.
165. Reeve BB, Wyrwich KW, Wu AW, Velikova G, Terwee CB, Snyder CF et al.: ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res* 2013, 22: 1889-1905.
166. Reinold MM, Gill TJ: Current concepts in the evaluation and treatment of the shoulder in overhead-throwing athletes, part 1: physical characteristics and clinical examination. *Sports Health* 2010, 2: 39-50.
167. Richards RR, An KN, Bigliani LU, Friedman RJ, Gartsman GM, Gristina AG et al.: A standardized method for the assessment of shoulder function. *J Shoulder Elbow Surg* 1994, 3: 347-352.
168. Roach KE, Budiman-Mak E, Songsiridej N, Lertratanakul Y: Development of a shoulder pain and disability index. *Arthritis Care Res* 1991, 4: 143-149.

169. Rocourt MH, Radlinger L, Kalberer F, Sanavi S, Schmid NS, Leunig M et al.: Evaluation of intratester and intertester reliability of the Constant-Murley shoulder assessment. *J Shoulder Elbow Surg* 2008, 17: 364-369.
170. Roe Y, Soberg HL, Bautz-Holter E, 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, 14: 73.
171. Romeo AA, Bach BR, Jr., O'Halloran KL: Scoring systems for shoulder conditions. *Am J Sports Med* 1996, 24: 472-476.
172. Rosales RS, Delgado EB, Diez de la Lastra-Bosch: 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.
173. Rosser R: The history of health related quality of life in 10 1/2 paragraphs. *J R Soc Med* 1993, 86: 315-318.
174. Rouleau DM, Faber K, Macdermid JC: Systematic review of patient-administered shoulder functional scores on instability. *J Shoulder Elbow Surg* 2010, 19: 1121-1128.
175. Roy JS, Macdermid JC, Woodhouse LJ: A systematic review of the psychometric properties of the Constant-Murley score. *J Shoulder Elbow Surg* 2010, 19: 157-164.
176. Saris-Baglama RN, Dewey CJ, Chisholm GB, Plumb E, King J, Rasicot P et al.: *QualityMetric Health Outcomes, Scoring Software 4.5, User's Guide*. 2011.
177. Schmidt S, Ferrer M, Gonzalez M, Gonzalez N, Valderas JM, Alonso J et al.: Evaluation of shoulder-specific patient-reported outcome measures: a systematic and standardized comparison of available evidence. *J Shoulder Elbow Surg* 2014, 23: 434-444.
178. Schmidt S, Garin O, Pardo Y, Valderas JM, Alonso J, Rebollo P et al.: Assessing quality of life in patients with prostate cancer: a systematic and standardized comparison of available instruments. *Qual Life Res* 2014, 23: 2169-2181.
179. Seidl H, Hunger M, Meisinger C, Kirchberger I, Kuch B, Leidl R et al.: The 3-Year Cost-Effectiveness of a Nurse-Based Case Management versus Usual Care for Elderly Patients with Myocardial Infarction: Results from the KORINNA Follow-Up Study. *Value Health* 2017, 20: 441-450.
180. Shrout PE, Fleiss JL: Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979, 86: 420-428.
181. Sijtsma K: Reliability Beyond Theory and Into Practice. *Psychometrika* 2009, 74: 169-173.
182. Singh JA, Sperling J, Buchbinder R, McMaken K: Surgery for shoulder osteoarthritis: A cochrane systematic review. *The Journal of Rheumatology* 2011, 38: 598-605.

183. Slobogean GP, Slobogean BL: Measuring shoulder injury function: common scales and checklists. *Injury* 2011, 42: 248-252.
184. Spearman C: General intelligence, objectively determined and measured. *The American Journal of Psychology* 1904, 15: 201-292.
185. Staples MP, Forbes A, Green S, Buchbinder R: Shoulder-specific disability measures showed acceptable construct validity and responsiveness. *J Clin Epidemiol* 2010, 63: 163-170.
186. Streiner DL: A checklist for evaluating the usefulness of rating scales. *Can J Psychiatry* 1993, 38: 140-148.
187. Tennant A, McKenna SP, Hagell P: Application of Rasch analysis in the development and application of quality of life instruments. *Value Health* 2004, 7 Suppl 1: S22-S26.
188. 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.
189. 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: 651-657.
190. Terwee CB, Prinsen CA, Ricci Garotti MG, Suman A, de Vet HC, Mokkink LB: The quality of systematic reviews of health-related outcome measurement instruments. *Qual Life Res* 2016, 25: 767-779.
191. The EuroQol Group: EuroQol--a new facility for the measurement of health-related quality of life. *Health Policy* 1990, 16: 199-208.
192. The WHOQOL Group: Study protocol for the World Health Organization project to develop a Quality of Life assessment instrument (WHOQOL). *Qual Life Res* 1993, 2: 153-159.
193. The WHOQOL Group: Development of the WHOQOL: Rationale and current status. *International Journal of Mental Health* 1994, 23: 24-56.
194. Torrens C, Orient F, Vila F, Escalada F, Marín M: Análisis comparativo entre la escala de Constant y el cuestionario de salud SF-36 en pacientes con patología subacromial. *Revista de Ortopedia y Traumatología* 2000, 44: 447-451.
195. Torres-Lacomba M, Sanchez-Sanchez B, Prieto-Gomez V, Pacheco-da-Costa S, Yuste-Sanchez MJ, Navarro-Brazalez B 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.
196. Uijen AA, Heinst CW, Schellevis FG, van den Bosch WJ, van de Laar FA, Terwee CB et al.: Measurement properties of questionnaires measuring continuity of care: a systematic review. *PLoS One* 2012, 7: e42256.
197. Valderas JM, Ferrer M, Alonso J: [Health-related quality of life instruments and other patient-reported outcomes]. *Med Clin (Barc)* 2005, 125 Suppl 1: 56-60.

198. Valderas JM, Ferrer M, Mendivil J, Garin O, Rajmil L, Herdman M et al.: Development of EMPRO: a tool for the standardized assessment of patient-reported outcome measures. *Value Health* 2008, 11: 700-708.
199. van de Water AT, Shields N, Davidson M, Evans M, Taylor NF: Reliability and validity of shoulder function outcome measures in people with a proximal humeral fracture. *Disabil Rehabil* 2014, 36: 1072-1079.
200. van den Ende CH, Rozing PM, Dijkmans BA, Verhoef JA, Voogt-van der Harst EM, Hazes JM: Assessment of shoulder function in rheumatoid arthritis. *J Rheumatol* 1996, 23: 2043-2048.
201. van der Heijden GJ: Shoulder disorders: a state-of-the-art review. *Baillieres Clin Rheumatol* 1999, 13: 287-309.
202. van der Heijden GJ, Leffers P, Bouter LM: Shoulder disability questionnaire design and responsiveness of a functional status measure. *J Clin Epidemiol* 2000, 53: 29-38.
203. Vergara I, Bilbao A, Orive M, Garcia-Gutierrez S, Navarro G, Quintana JM: Validation of the Spanish version of the Lawton IADL Scale for its application in elderly people. *Health Qual Life Outcomes* 2012, 10: 130.
204. Vicente-Herrero MT, Capdevilla-Garcia L, Lopez Gonzalez L, Ramirez-Iñiguez de la Torre MV (Eds): *El hombro y sus patologías en medicina del trabajo*. In SEMERGEN 2009, 35: 197-202.
205. Vilagut G, Ferrer M, Rajmil L, Rebollo P, Permanyer-Miralda G, Quintana JM et al.: [El Cuestionario de Salud SF-36 español: una década de experiencia y nuevos desarrollos]. *Gac Sanit* 2005, 19: 135-150.
206. Vincent HK, Struk AM, Reed A, Wright TW: Mid-term shoulder functional and quality of life outcomes after shoulder replacement in obese patients. *Springerplus* 2016, 5: 1929.
207. Virta L, Joranger P, Brox JI, Eriksson R: Costs of shoulder pain and resource use in primary health care: a cost-of-illness study in Sweden. *BMC Musculoskelet Disord* 2012, 13: 17.
208. Vrotsou K, Cuellar R, Silio F, Rodriguez MA, Garay D, Busto G et al.: Patient self-report section of the ASES questionnaire: a Spanish validation study using classical test theory and the Rasch model. *Health Qual Life Outcomes* 2016, 14: 147.
209. Walker GA: *Common Statistical Methods for Clinical Research with SAS Examples*, Second Edition edn. SAS Publishing; 2002.
210. Walton MJ, Walton JC, Honorez LA, Harding VF, Wallace WA: A comparison of methods for shoulder strength assessment and analysis of Constant score change in patients aged over fifty years in the United Kingdom. *J Shoulder Elbow Surg* 2007, 16: 285-289.
211. Ware J, Jr., Kosinski M, Keller SD: A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care* 1996, 34: 220-233.

212. Ware JE, Jr., Kemp JP, Buchner DA, Singer AE, Nolop KB, Goss TF: The responsiveness of disease-specific and generic health measures to changes in the severity of asthma among adults. *Qual Life Res* 1998, 7: 235-244.
213. Ware JE, Jr., Kosinski M, Dewey JE, Gandek B. How to score and interpret single-item health status measures: a manual for users of the SF-8 Health Survey. Lincoln RI: Quality Metric Incorporated; 2001.
214. Ware JE, Jr., Sherbourne CD: The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992, 30: 473-483.
215. Weng LJ: Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. *Educational and Psychological measurement* 2004, 64: 956-972.
216. World Health Organization. The World Health Organization Quality of Life. 2017. 18-3-2017.
217. Wright B.D. Model selection: Rating Scale Model (RSM) or Partial Credit Model (PCM)? *Rasch Measurement Transactions* 12[3], 641-642. 1998.
218. Wright BD, Linacre JM. Reasonable mean-square fit values. *Rasch Measurement Transactions* 8[3], 370. 1994.
219. Wright RW, Baumgarten KM: Shoulder outcomes measures. *J Am Acad Orthop Surg* 2010, 18: 436-444.
220. Wylie JD, Bershadsky B, Iannotti JP: The effect of medical comorbidity on self-reported shoulder-specific health related quality of life in patients with shoulder disease. *J Shoulder Elbow Surg* 2010, 19: 823-828.
221. Yahia A, Guermazi M, Khmekhem M, Ghroubi S, Ayedi K, Elleuch MH: Translation into Arabic and validation of the ASES index in assessment of shoulder disabilities. *Ann Phys Rehabil Med* 2011, 54: 59-72.
222. Yian EH, Ramappa AJ, Arneberg O, Gerber C: The Constant score in normal shoulders. *J Shoulder Elbow Surg* 2005, 14: 128-133.
223. Zumbo BD, Chan EKH. *Validity and Validation in Social, Behavioural, and Health Sciences*. 2014. Springer International Publishing. Social Indicators Research Series.

## Anexos

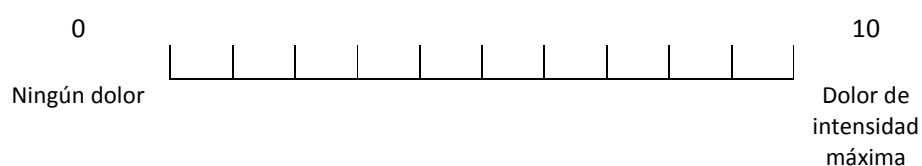
### Anexo I: Cuestionarios y Escalas Utilizadas

#### Versión castellana de la escala ASES-p

##### Auto-evaluación del hombro afectado

Dolor:

¿Qué intensidad de dolor siente hoy? (haga una marca en la línea)



Función: Rodee con un círculo el número que describa su capacidad para realizar las siguientes actividades:

0 = Imposible; 1= Con mucha dificultad; 2=Con cierta dificultad; 3=Sin ninguna dificultad

Actividad	Con brazo derecho				Con brazo izquierdo			
	0	1	2	3	0	1	2	3
1. Ponerse un abrigo	0	1	2	3	0	1	2	3
2. Dormir sobre el lado afectado o dolorido	0	1	2	3	0	1	2	3
3. Lavarse la espalda/abrocharse el sujetador por detrás	0	1	2	3	0	1	2	3
4. Limpiarse tras la micción y defecación	0	1	2	3	0	1	2	3
5. Peinarse	0	1	2	3	0	1	2	3
6. Llegar a una balda alta	0	1	2	3	0	1	2	3
7. Levantar 4,5kg por encima del hombro (i.e. 3 botellas de agua de 1,5kg)	0	1	2	3	0	1	2	3
8. Lanzar una pelota por encima de la cabeza (i.e. saque de banda, tirar una piedra)	0	1	2	3	0	1	2	3
9. Realizar su trabajo habitual- Lista: .....	0	1	2	3	0	1	2	3
10. Realizar su deporte habitual- Lista: .....	0	1	2	3	0	1	2	3

# **Índice de Barthel (actividades básicas de la vida diaria)**

**Nota: Marque con una cruz la opción que mejor le describa en este momento. En caso de que necesite la ayuda de otra persona para hacer alguna de estas actividades, se marca como 'Dependiente'.**

## **¿Necesita ayuda para**

### **1. Alimentarse**

- Dependiente: necesita ser alimentado por otra persona.
- Necesita ayuda: por ejemplo, para cortar, extender la mantequilla, etc., pero es capaz de comer solo.
- Independiente: Capaz de comer solo en un tiempo razonable. La comida puede ser cocinada y servida por otra persona.

### **2. El baño (lavarse)**

- Dependiente: necesita alguna ayuda.
- Independiente: Capaz de lavarse entero/a, de entrar y salir del baño sin ayuda y de hacerlo sin que una persona le supervise.

### **3. Vestirse**

- Dependiente: incapaz de manejarse sin asistencia mayor.
- Necesita ayuda: pero hace al menos la mitad de las tareas en un tiempo razonable.
- Independiente: Es capaz de ponerse, quitarse la ropa sin ayuda. Se ata los zapatos, abrocha los botones, etc. Se coloca el braguero o el corsé si lo precisa.

### **4. Asearse (lavarse las manos, la cara, peinarse, etc.)**

- Dependiente: necesita alguna ayuda.
- Independiente: realiza todas las tareas personales sin ayuda alguna (lavarse las manos, la cara, peinarse, etc.). Incluye afeitarse y lavarse los dientes. Incluye manejar el enchufe si la maquinilla es eléctrica: los complementos necesarios pueden ser provistos por alguna persona.

### **5. Deposición (Evacuar) A la hora de 'hacer de cuerpo' ¿Tiene problemas para llegar al servicio a tiempo?**

- Incontinente (más de una vez a la semana)
- Accidente ocasional: raro (menos de una vez por semana), o necesita ayuda para el enema o los supositorios.
- Continente, ningún accidente: si necesita enema o supositorios se arregla por sí solo.

### **6. Micción (Orinar) ¿y para orinar?**

- Incontinente (más de un episodio en 24h)
- Accidente ocasional: como máximo una vez al día o necesita ayuda con la sonda, orinal, botella etc.
- Continente, ningún accidente: seco día y noche. Capaz de usar cualquier dispositivo (sonda, orinal, cambiar la bolsa) por sí solo.

### **7. Ir al retrete (ir al servicio)**

- Dependiente: incapaz de acceder al retrete o de utilizarlo sin asistencia mayor.
- Necesita ayuda: capaz de manejarse con una pequeña ayuda; es capaz de usar el cuarto de baño. Puede limpiarse solo.
- Independiente: entra y sale solo y no necesita ayuda alguna por parte de otra persona.



### **8. Trasladarse del sillón a la cama**

- Dependiente: necesita grúa o alzamiento completo por dos personas. Incapaz de permanecer sentado.
- Gran ayuda: capaz de estar sentado sin ayuda, pero necesita mucha asistencia para entrar o salir de la cama.
- Mínima ayuda: incluye supervisión verbal o pequeña ayuda física.
- Independiente: no necesita ayuda. Si utiliza silla de ruedas, lo hace independientemente.

### **9. Andar, moverse caminando (deambular)**

- Dependiente: requiere ayuda mayor.
- Independiente en silla de ruedas: no requiere ayuda ni supervisión.
- Necesita ayuda: necesita supervisión física o verbal por parte de otra persona o utiliza andador.
- Independiente: puede usar cualquier ayuda (prótesis, bastones, muletas, etc.), excepto andador. La velocidad no es importante. Puede caminar al menos 50 m o equivalente sin ayuda o supervisión.

### **10. Subir y bajar escalones**

- Dependiente: es incapaz. Necesita usar el ascensor.
- Necesita ayuda: supervisión física o verbal.
- Independiente: capaz de subir y bajar un piso de escaleras sin ayuda o supervisión, aunque utilice barandilla o instrumentos de apoyo.

## Escala SF-36

Las preguntas que siguen se refieren a lo que usted piensa sobre su salud. Sus respuestas permitirán saber cómo se encuentra usted y hasta qué punto es capaz de hacer sus actividades habituales. *¡Gracias por contestar a estas preguntas!*

Para cada una de las siguientes preguntas, por favor marque con una  la casilla que mejor corresponda a su respuesta.

1. En general, usted diría que su salud es:

Excelente	Muy buena	Buena	Regular	Mala
▼	▼	▼	▼	▼
<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5

2. ¿Cómo diría usted que es su salud actual, comparada con la de hace un año?

Mucho mejor ahora que hace un año	Algo mejor ahora que hace un año	Más o menos igual que hace un año	Algo peor ahora que hace un año	Mucho peor ahora que hace un año
▼	▼	▼	▼	▼
<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5

**3. Las siguientes preguntas se refieren a actividades o cosas que usted podría hacer en un día normal. Su salud actual, ¿le limita para hacer esas actividades o cosas? Si es así, ¿cuánto?**

Sí, me limita mucho	Sí, me limita un poco	No, no me limita nada
▼	▼	▼

- a Esfuerzos intensos, tales como correr, levantar objetos pesados, o participar en deportes agotadores.....  1 .....  2 .....  3
- b Esfuerzos moderados, como mover una mesa, pasar la aspiradora, jugar a los bolos o caminar más de 1 hora .....  1 .....  2 .....  3
- c Coger o llevar la bolsa de la compra .....  1 .....  2 .....  3
- d Subir varios pisos por la escalera .....  1 .....  2 .....  3
- e Subir un solo piso por la escalera .....  1 .....  2 .....  3
- f Agacharse o arrodillarse .....  1 .....  2 .....  3
- g Caminar un kilómetro o más .....  1 .....  2 .....  3
- h Caminar varios centenares de metros .....  1 .....  2 .....  3
- i Caminar unos 100 metros .....  1 .....  2 .....  3
- j Bañarse o vestirse por sí mismo .....  1 .....  2 .....  3

**4. Durante las 4 últimas semanas, ¿con qué frecuencia ha tenido alguno de los siguientes problemas en su trabajo u otras actividades cotidianas a causa de su salud física?**

Siempre	Casi siempre	Algunas veces	Casi nunca	Nunca
▼	▼	▼	▼	▼

- a ¿Tuvo que reducir el tiempo dedicado al trabajo u otras actividades cotidianas? .....  1 .....  2 .....  3 .....  4 .....  5
- b ¿Hizo menos de lo que hubiera querido hacer? .....  1 .....  2 .....  3 .....  4 .....  5
- c ¿Estuvo limitado en el tipo de trabajo u otras actividades? .....  1 .....  2 .....  3 .....  4 .....  5
- d ¿Tuvo dificultad para hacer su trabajo u otras actividades cotidianas (por ejemplo, le costó más de lo normal)? .....  1 .....  2 .....  3 .....  4 .....  5

**5. Durante las 4 últimas semanas, ¿con qué frecuencia ha tenido alguno de los siguientes problemas en su trabajo u otras actividades cotidianas a causa de algún problema emocional (como estar triste, deprimido o nervioso)?**

Siempre	Casi siempre	Algunas veces	Casi nunca	Nunca
▼	▼	▼	▼	▼

- a ¿Tuvo que reducir el tiempo dedicado al trabajo u otras actividades cotidianas? .....  1 .....  2 .....  3 .....  4 .....  5
- b ¿Hizo menos de lo que hubiera querido hacer? .....  1 .....  2 .....  3 .....  4 .....  5
- c ¿Hizo su trabajo u otras actividades cotidianas menos cuidadosamente que de costumbre? .....  1 .....  2 .....  3 .....  4 .....  5

6. Durante las 4 últimas semanas, ¿hasta qué punto su salud física o los problemas emocionales han dificultado sus actividades sociales habituales con la familia, los amigos, los vecinos u otras personas?

Nada	Un poco	Regular	Bastante	Muchísimo
▼	▼	▼	▼	▼
<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5

7. ¿Tuvo dolor en alguna parte del cuerpo durante las 4 últimas semanas?

No, ninguno	Sí, muy poco	Sí, un poco	Sí, moderado	Sí, mucho	Sí, muchísimo
▼	▼	▼	▼	▼	▼
<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6

8. Durante las 4 últimas semanas, ¿hasta qué punto el dolor le ha dificultado su trabajo habitual (incluido el trabajo fuera de casa y las tareas domésticas)?

Nada	Un poco	Regular	Bastante	Muchísimo
▼	▼	▼	▼	▼
<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5

9. Las preguntas que siguen se refieren a cómo se ha sentido y cómo le han ido las cosas durante las 4 últimas semanas. En cada pregunta responda lo que se parezca más a cómo se ha sentido usted. Durante las últimas 4 semanas ¿con qué frecuencia...

	Siempre	Casi siempre	Algunas veces	Casi nunca	Nunca
a se sintió lleno de vitalidad? .....	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
b estuvo muy nervioso? .....	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
c se sintió tan bajo de moral que nada podía animarle?.....	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
d se sintió calmado y tranquilo? .....	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
e tuvo mucha energía?.....	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
f se sintió desanimado y deprimido?.....	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
g se sintió agotado? .....	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
h se sintió feliz?.....	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
i se sintió cansado? .....	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5

10. Durante las 4 últimas semanas, ¿con qué frecuencia la salud física o los problemas emocionales le han dificultado sus actividades sociales (como visitar a los amigos o familiares)?

Siempre	Casi siempre	Algunas veces	Casi nunca	Nunca
<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5

11. Por favor, diga si le parece **CIERTA** o **FALSA** cada una de las siguientes frases:

Totalmente cierta	Bastante cierta	No lo sé	Bastante falsa	Totalmente falsa
▼	▼	▼	▼	▼

- a Creo que me pongo enfermo más fácilmente que otras personas .....  1 .....  2 .....  3 .....  4 .....  5
- b Estoy tan sano como cualquiera .....  1 .....  2 .....  3 .....  4 .....  5
- c Creo que mi salud va a empeorar .....  1 .....  2 .....  3 .....  4 .....  5
- d Mi salud es excelente .....  1 .....  2 .....  3 .....  4 .....  5

***¡Gracias por contestar a estas preguntas!***

# Cuestionario de Constant-Murley Score

DATOS DEL PACIENTE
--------------------

<b>DIAGNÓSTICO:</b>	<b>LADO:</b> Derecho Izquierdo
<b>EXPLORACIÓN:</b> PRE-OP	3 MESES 6 MESES
<b>FECHA REVISIÓN</b>	_____

## A.- DOLOR (máximo 15 puntos)

### 1. ¿Tiene usted dolor en su hombro afectado? (cuando realiza actividades normales):

No = 15 pts      Leve= 10 pts      Moderado= 5 pts      Severo o Permanente= 0 pts

### 2. Escala lineal:

Si "0" indica No dolor y "15" es el *máximo dolor que usted ha llegado a sufrir*, por favor indique cuál es su nivel de dolor en el hombro afectado. Los puntos se asignan por los investigadores

<b>Nivel de Dolor:</b>	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>11</b>	<b>12</b>	<b>13</b>	<b>14</b>	<b>15</b>
<i>Puntos</i>	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0

## B.- ACTIVIDAD VIDA DIARIA (máximo 20 puntos)

### 1. ¿Su trabajo o vida diaria está limitada por su hombro afectado?

No = 4,      Limitación Moderada = 2,      Limitación Severa = 0

### 2. ¿Sus actividades de ocio están limitadas por su hombro afectado?

No = 4,      Limitación Moderada = 2,      Limitación Severa = 0

### 3. ¿Tiene problemas de sueño durante la noche debido a su hombro afectado?

No = 2,      A veces = 1,      Sí = 0

### 4. Indique hasta qué altura puede usar su brazo para realizar actividades normales sin experimentar dolor (limitación en posición)

Cintura = 2,      Xifoides = 4,      Cuello = 6,      Cabeza = 8,      Por encima de la cabeza = 10

## C.- MOVILIDAD (máximo 40 puntos)

<b>1. Flexión en grados</b>	0- 30 0 pts	31 - 60 2 pts	61- 90 4 pts	91- 120 6 pts	121- 150 8 pts	>150 10 pts
<b>2. Abducción en grados</b>	0- 30 0 pts	31 - 60 2 pts	61- 90 4 pts	91- 120 6 pts	121- 150 8 pts	>150 10 pts
<b>3. Rotación externa (se pregunta todo)</b>	Mano detrás de la cabeza y codo delante -2 pts	Mano detrás de la cabeza y codo detrás - 4 pts	Mano encima de la cabeza y codo delante - 6 pts	Mano encima de la cabeza y codo detrás -8 pts	Elevación completa de brazo - 10 pts	
<b>4. Rotación interna dorso a mano (se pregunta todo)</b>	Muslo 0 pts	Glúteo 2 pts	Sacro-ílica 4 pts	Cintura 6 pts	D12 8pts	Interscapular 10pts

## D.- FUERZA MUSCULAR (máximo 25 puntos)

### 1. ¿Cuál es el máximo peso que puede levantar el paciente durante 5 segundos tres veces consecutivas en los 5 intentos? Si el no fuese capaz de mantener el peso en 90º abducción, anotar el peso y los grados en los que sí puede.

<b>Peso levantado en 90º de abducción</b>	<b>0 kg</b>	<b>2.5 kg</b>	<b>5kg</b>	<b>7.5kg</b>	<b>10kg</b>	<b>12.5kg</b>
Peso levantado en ____º de abducción	0 kg	2.5 kg	5kg	7.5kg	10kg	12.5kg



## EMPRO

Cuestionario para la evaluación de instrumentos de medida  
de resultados percibidos por los pacientes  
(versión: junio 2007)

© 2007

**INSTRUMENTO:**

**EVALUADOR:**

**IMPORTANTE:**

Por favor, lea detenidamente las instrucciones antes de utilizar el instrumento EMPRO.

Proyecto financiado por: FONDO DE INVESTIGACIÓN SANITARIA- ISCIII: [Expte. P106/90483](#)

### INTRODUCCIÓN

#### ¿Qué es EMPRO y para qué sirve?

- La idea de EMPRO nació dentro de la Red IRYSS
- EMPRO es una herramienta para la evaluación estandarizada de los instrumentos de resultados percibidos en los pacientes que se utilizan en la Investigación en Resultados de Salud y Servicios Sanitarios
- Consta de un cuestionario -que rellena cada evaluador de forma individual- sobre las características del instrumento a evaluar así como un **procedimiento de consenso** que pretende ayudar a los evaluadores llegar a un acuerdo sobre el instrumento evaluado
- El **contenido** de EMPRO se ha basado fundamentalmente en las recomendaciones incluidas en el artículo del *Scientific Advisory Committee of the Medical Outcomes Trust*<sup>1</sup> sobre las características métricas que idealmente debe tener un instrumento de medida de resultados percibidos por los pacientes
- Para definir los **procedimientos** utilizados en el proceso de evaluación se ha seguido el modelo de la colaboración AGREE<sup>2</sup> para la evaluación de guías de práctica clínica.

#### ¿Qué instrumentos se pueden evaluar con EMPRO?

- EMPRO está diseñado para evaluar herramientas de medida de resultados centrados en los pacientes, preferentemente los centrados en la calidad de vida relacionada con la salud.
- Puede aplicarse a instrumentos desarrollados para cualquier tipo de patología y cualquier forma de administración (por ejemplo, instrumentos desarrollados en papel, en soporte electrónico...)
- Se puede utilizar tanto para las versiones originales de los instrumentos como para las versiones adaptadas culturalmente (en el primer caso, se obvia la sección 8 sobre adaptación cultural de los instrumentos).

<sup>1</sup>Scientific Advisory Committee of the Medical Outcomes Trust. Assessing health status and quality-of-life instruments: Attributes and review criteria. *Quality of Life Research*. 2002; 11: 193-205.

<sup>2</sup>AGREE Collaboration. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. *Qual Saf Health Care* 2003;12:18-23.

## EL PROCESO DE EVALUACIÓN

### 1. Estructura y contenidos de EMPRO

EMPRO consta de 39 ítems organizados en ocho áreas. Cada área abarca una dimensión diferenciada de la calidad del instrumento a evaluar.

#### I. CONCEPTO Y MODELO DE MEDIDA (ítems 1-7).

Se refiere a la justificación y descripción del concepto y de las poblaciones que una medida intenta evaluar y la relación entre ambos.

#### II. FIABILIDAD (ítems 8-15).

Se refiere al grado en que un instrumento no tiene error aleatorio.

#### III VALIDEZ (ítems 16-21).

Se refiere a la capacidad del instrumento de medir lo que pretende medir.

#### IV. SENSIBILIDAD AL CAMBIO (ítems 22-24).

Se refiere a la habilidad de un instrumento para detectar cambios en el tiempo.

#### V. INTERPRETABILIDAD (ítems 25-27).

Se refiere a la facilidad para asignar un significado comprensible a las puntuaciones del instrumento.

#### VI. FACILIDAD DE USO DEL INSTRUMENTO (ítems 28-34).

Se refiere al tiempo, esfuerzo y otros requisitos que se requieren para administrar el instrumento. Incluye a la persona a quien se administra el instrumento (carga del entrevistado) y a quienes lo administran (carga administrativa).

#### VII. FORMAS ALTERNATIVAS DE ADMINISTRACIÓN (ítems 35-36).

Se refiere a la utilización de formas alternativas de administración del instrumento.

#### VIII. PROCESO DE ADAPTACIÓN CULTURAL Y LINGÜÍSTICA (ítems 37-39).

Se refiere a la adaptación cultural y lingüística que se ha realizado del instrumento.

### 2. Los evaluadores

Cada instrumento será evaluado por al menos dos evaluadores. Los evaluadores deben ser expertos en el desarrollo, validación y/o uso de los instrumentos de resultados centrados en los pacientes.

### 3. El proceso de consenso

Los desacuerdos entre evaluadores serán comentados por los evaluadores para llegar a un consenso y en caso de que persistan desacuerdos se procederá a una tercera evaluación.

### 4. Documentación

Para evaluar un instrumento, los evaluadores dispondrán de la información relativa a su desarrollo y aplicación. Esta información puede estar incluida en el mismo instrumento, en su manual, en artículos científicos o en otras publicaciones. Recomendamos la lectura del instrumento y de toda la documentación relacionada antes de comenzar la evaluación. Toda la documentación entregada a los evaluadores será numerada para facilitar su identificación en los comentarios.

### 5. Método de puntuación

Se calculan puntuaciones para cada área de EMPRO teniendo en cuenta las puntuaciones de cada evaluador. EMPRO también proporciona una puntuación global. En el cálculo de las puntuaciones se tienen en cuenta tanto los ítems para los cuales se considera que falta información como los ítems en los cuales los evaluadores han marcado 'No aplicable'.

## INSTRUCCIONES PARA EL EVALUADOR

### 1. Información sobre los criterios a evaluar

Para cada área e ítem del instrumento, se ha elaborado información adicional de carácter orientativo. Por favor, lea esta información detalladamente antes de dar su respuesta.

### 2. Escala de respuesta

Cada ítem se responde en una escala con 4 opciones de respuesta (desde "Muy de acuerdo" hasta "Muy en desacuerdo"). El evaluador debe utilizar esta escala para dar su opinión sobre el grado en que el instrumento cumple con los criterios señalados.

Muy de acuerdo  4     3     2     1    No aplicable

Muy en desacuerdo    No hay información

En todos los ítems, también existe una opción de 'No hay información' que el evaluador debe señalar si cree que no existe la información suficiente y necesaria en el material proporcionado para poder hacer una evaluación.

Excepcionalmente, existe la opción de 'No aplicable' que el evaluador debe señalar si cree que el ítem no es susceptible de ser aplicado al cuestionario que se está evaluando.

### 3. Justificación y comentarios

Después de la escala de respuesta de cada ítem, el evaluador encontrará el siguiente recuadro:

Referencia:	Comentarios:
-------------	--------------

En el cuadro de la izquierda podrá indicar en qué documento encontró la información en la que se basa su evaluación. Para ello deberá indicar el número de documento, según de la página entre paréntesis. Se podrán referenciar cuantos documentos sean necesarios. En el cuadro de la derecha, se podrán introducir los comentarios que se consideren pertinentes. Por ejemplo, el evaluador puede querer indicar que está "Muy en desacuerdo" con el ítem porque la metodología descrita en la información proporcionada no es satisfactoria. Al final del instrumento hay un espacio donde se pueden introducir comentarios de carácter más general.

### 4. Evaluación general

Al final del documento se incluye un apartado para realizar una evaluación general del instrumento teniendo en cuenta cada uno de los criterios de evaluación. También se debe especificar el uso recomendado, ya sea "Uso longitudinal" o "Uso Transversal". Asimismo, deberá recomendar el tipo de población en la que se aplica el instrumento o el ámbito del mismo así como el tipo de estudio y las posibles evaluaciones que se pueden realizar con el instrumento.

## I. CONCEPTO Y MODELO DE MEDIDA

**El concepto que una medida intenta evaluar, las poblaciones a las que se quiere aplicar y la relación entre ambos conceptos deben estar adecuadamente argumentados.**

**Marque una sola casilla**

1. El concepto que se pretende medir está suficientemente descrito.

Se evaluarán los siguientes aspectos:

- Se indica el concepto general que el instrumento pretende evaluar (por ejemplo estado funcional, bienestar, calidad de vida relacionada con la salud, satisfacción, etc.).
- Se proporciona una lista de todas las dimensiones, en el supuesto de que exista más de una dimensión.

Muy de acuerdo  4  3  2  1  Muy en desacuerdo  No hay información

Referencia:

Comentarios:

2. Las bases conceptuales y empíricas para la obtención de los ítems del instrumento y combinarlos en una o diversas dimensiones están suficientemente descritas y son adecuadas.

Se evaluarán los siguientes aspectos:

- Las bases conceptuales y los métodos para la obtención de los ítems (ej. revisión de la literatura, grupos de expertos, otros cuestionarios, etc.)
- Se expone con claridad por qué se combinan diferentes ítems en una o varias dimensiones (ej. análisis factorial, grupos de expertos).
- Se expone con claridad por qué un ítem se incluye en una dimensión y no en otra.

Muy de acuerdo  4  3  2  1  Muy en desacuerdo  No hay información

Referencia:

Comentarios:

### RECUERDE

Se valorará si se aporta suficiente información, los métodos son adecuados y los resultados obtenidos indican buen funcionamiento del instrumento. Se puntuará como 4 o 3 si se cumplen todos o la mayoría de los criterios indicados. En caso contrario, la puntuación no debería superar el 2.

<p><b>3. La dimensionalidad y la distinción de escalas están suficientemente descritas y razonadas.</b></p> <p>Se evaluarán los siguientes aspectos:</p> <ul style="list-style-type: none"> <li>- Se fundamenta la existencia de diferentes dimensiones o de una única dimensión</li> <li>- Se define el constructo medido en cada dimensión.</li> <li>- Se justifica la construcción de un índice, en el supuesto de que la tenga.</li> </ul>	<p>4   3   2   1</p> <p>Muy de acuerdo   Muy en desacuerdo</p>	<p>No hay información</p>
Referencia:	Comentarios:	
<p><b>4. La participación de la población diana en la obtención del contenido del instrumento está suficientemente descrita, con métodos adecuados y resultados satisfactorios.</b></p> <p>Se evaluarán los siguientes aspectos:</p> <ul style="list-style-type: none"> <li>- La población diana participa en la obtención del pool de ítems (ej.: grupos focales).</li> <li>- Se evalúa la adecuación del contenido del instrumento para la población diana (ej.: prueba piloto, método del impacto clínico).</li> </ul>	<p>4   3   2   1</p> <p>Muy de acuerdo   Muy en desacuerdo</p>	<p>No hay información</p>
Referencia:	Comentarios:	
<p><b>5. La variabilidad de la escala en la población está suficientemente descrita y es adecuada para el propósito del instrumento.</b></p> <p>Se evaluarán los siguientes aspectos:</p> <ul style="list-style-type: none"> <li>- La distribución, la tendencia central y la dispersión de las puntuaciones.</li> <li>- Los efectos techo y suelo.</li> <li>- El patrón de los valores peraltados de la escala.</li> </ul>	<p>4   3   2   1</p> <p>Muy de acuerdo   Muy en desacuerdo</p>	<p>No hay información</p>
Referencia:	Comentarios:	
<p><b>6. La escala de medida está suficientemente descrita y razonada.</b></p> <p>Se evaluarán los siguientes aspectos:</p>		
RECUERDE	Se valorará si se aporta suficiente información, los métodos son adecuados y los resultados obtenidos indican buen funcionamiento del instrumento. Se puntuará como 4 o 3 si se cumplen todos o la mayoría de los criterios indicados. En caso contrario, la puntuación no debería superar el 2.	

<p>- Se especifica la escala de medida (ej. ordinal, de intervalo, de razón).</p> <p>- Se aporta evidencia empírica que justifique este nivel de medida de la escala.</p>	<p>4   3   2   1</p> <p>Muy de acuerdo   Muy en desacuerdo</p>	<p>No hay información</p>
Referencia:	Comentarios:	
<p><b>7. La justificación y el método para el cálculo de las puntuaciones (global y/o por dimensiones) a partir de las puntuaciones brutas están suficientemente descritos.</b></p> <p>Se evaluarán los siguientes aspectos:</p> <ul style="list-style-type: none"> <li>- Se describen los procedimientos y transformaciones para obtener la puntuación final de cada una de las dimensiones a partir de las puntuaciones crudas.</li> <li>- Se argumenta el hecho de ponderar o no ciertos ítems o dimensiones.</li> </ul>	<p>4   3   2   1</p> <p>Muy de acuerdo   Muy en desacuerdo</p>	<p>No hay información</p>
Referencia:	Comentarios:	
RECUERDE	Se valorará si se aporta suficiente información, los métodos son adecuados y los resultados obtenidos indican buen funcionamiento del instrumento. Se puntuará como 4 o 3 si se cumplen todos o la mayoría de los criterios indicados. En caso contrario, la puntuación no debería superar el 2.	

## II. FIABILIDAD

### II.1. Consistencia interna.

Es una estimación de la precisión del instrumento basada en la homogeneidad de los ítems en una administración. Su indicador en psicometría clásica es el coeficiente alfa de Cronbach. El coeficiente de Kuder-Richardson KR-20 es apropiado para opciones de respuesta dicotómicas. El error estándar de medida (EEM) se calcula multiplicando la desviación estándar de la puntuación por la raíz cuadrada del complementario de la fiabilidad (1-fiabilidad). En la aproximación de la Teoría de Respuesta al Ítem (TRI) se debe evaluar la precisión a lo largo de todo el rango de la escala. También se puede calcular la fiabilidad marginal, valor único análogo a la aproximación clásica.

8. Los métodos empleados para obtener los datos de consistencia interna están suficientemente descritos y son adecuados.

Se evaluarán los siguientes aspectos:

- Descripción del cálculo del tamaño y obtención de la muestra.
- Características de la muestra (socio-demográficas, clínicas si proviene de una población de pacientes, etc.).
- Las condiciones de administración (cómo y dónde fue administrado el instrumento)
- Estadística descriptiva de los resultados obtenidos con el instrumento (medias, desviación estándar, efectos suelo y techo).

Muy de acuerdo  4  3  2  1  Muy en desacuerdo  No hay información

Referencia:

Comentarios:

9. Los valores de los coeficientes Alfa de Cronbach y/o KR-20 son adecuados.

Se puntuará como 4 cuando los coeficientes sean  $>0,9$  para la/los puntuación/es globales y  $>0,7$  en todas las dimensiones. Si la mayoría de los coeficientes de las puntuaciones globales o de las dimensiones no supera el 0,7 la puntuación no podrá ser superior a 1.

Muy de acuerdo  4  3  2  1  Muy en desacuerdo  No hay información

Referencia:

Comentarios:

RECUERDE

Se valorará si se aporta suficiente información, los métodos son adecuados y los resultados obtenidos indican buen funcionamiento del instrumento. Se puntuará como 4 o 3 si se cumplen todos o la mayoría de los criterios indicados. En caso contrario, la puntuación no debería superar el 2.

10. La estimación de la fiabilidad según la aproximación Teoría de Respuesta a los Ítems (TRI) está suficientemente descrita y es adecuada.

Se evaluarán los siguientes aspectos:

- La gráfica con los EEM (error estándar de la medida) a lo largo del rango de la escala
- La fiabilidad marginal de la aproximación TRI para cada una de las puntuaciones.

Muy de acuerdo  4  3  2  1  Muy en desacuerdo  No hay información

Referencia:

Comentarios:

11. La estimación de la consistencia interna en cada población de interés está suficientemente descrita.

Si existen razones para pensar que la consistencia interna puede variar en determinadas grupos o poblaciones (por ejemplo diferentes patologías, grupos culturales, grupos de edades), se deben aportar datos para cada uno de ellos. Si no se aporta ninguna información porque no existen razones para considerar tales diferencias entre poblaciones se puntuará como no aplicable.

Muy de acuerdo  4  3  2  1  Muy en desacuerdo  No aplicable  No hay información

Referencia:

Comentarios:

RECUERDE

Se valorará si se aporta suficiente información, los métodos son adecuados y los resultados obtenidos indican buen funcionamiento del instrumento. Se puntuará como 4 o 3 si se cumplen todos o la mayoría de los criterios indicados. En caso contrario, la puntuación no debería superar el 2.

**11.2. Reproducibilidad.**

Incluye la estabilidad del instrumento a lo largo del tiempo (test-retest en entrevistados que se asume que no han cambiado) y concordancia entre entrevistadores en un mismo momento (para instrumentos administrados por entrevistador). Su indicador por excelencia en la psicometría clásica es el coeficiente de correlación intraclass (CCI) para variables de intervalo, y la Kappa simple y ponderada para escalas nominales y ordinales, respectivamente. En la aproximación de la Teoría de Respuesta al ítem (TRI) la estabilidad se evalúa en diferentes puntos a lo largo de la escala.

12. Los métodos empleados para obtener los datos de reproducibilidad están suficientemente descritos y son adecuados.

Se evaluarán los siguientes aspectos:

- Descripción del cálculo del tamaño y el método de obtención de la muestra.
- Características de la muestra (socio-demográficas, clínicas si proviene de una población de pacientes, etc.).
- Las condiciones de administración (cómo y dónde fue administrado el instrumento).
- Estadística descriptiva de los resultados obtenidos con el instrumento.

Muy de acuerdo  4  3  2  1  Muy en desacuerdo  No hay información

Referencia:

Comentarios:

13. El diseño de la comparación test-retest y el intervalo entre administraciones están suficientemente descritos y razonados y son adecuados.

Se evaluarán los siguientes aspectos:

- El intervalo de tiempo entre administraciones.
- El método empleado para determinar la estabilidad de la muestra (pregunta de cambio de salud percibida u otros instrumentos genéricos o específicos).
- La independencia entre entrevistadores.

Muy de acuerdo  4  3  2  1  Muy en desacuerdo  No hay información

Referencia:

Comentarios:

**RECUERDE**

Se valorará si se aporta suficiente información, los métodos son adecuados y los resultados obtenidos indican buen funcionamiento del instrumento. Se puntuará como 4 o 3 si se cumplen todos o la mayoría de los criterios indicados. En caso contrario, la puntuación no debería superar el 2.

14. Los valores del coeficiente de fiabilidad de la prueba test-retest y/o la concordancia entre entrevistadores para todas las puntuaciones están suficientemente descritos y son adecuados.

Se puntuará como 4 cuando el CCI sea  $>0.9$  para la/las puntuación/es globales y  $>0.7$  para todas las dimensiones. Si la mayoría de puntuaciones de cualquier tipo no supera el 0.7 la puntuación tendría que ser 1. No existen valores de referencia para la estimación mediante coeficiente Kappa. Cuando se aporten estimadores distintos, como la media de la diferencia, correlaciones, límites de la concordancia o curvas ROC, como máximo se podría puntuar con un 2.

Muy de acuerdo  4  3  2  1  Muy en desacuerdo  No hay información

Referencia:

Comentarios:

15. La estimación de la concordancia según aproximación TRI está suficientemente descrita y es adecuada.

Se evaluarán los siguientes aspectos:

- La comparabilidad de los parámetros estimados de los ítems.
- La precisión de la medida en administraciones repetidas.

Muy de acuerdo  4  3  2  1  Muy en desacuerdo  No hay información

Referencia:

Comentarios:

**RECUERDE**

Se valorará si se aporta suficiente información, los métodos son adecuados y los resultados obtenidos indican buen funcionamiento del instrumento. Se puntuará como 4 o 3 si se cumplen todos o la mayoría de los criterios indicados. En caso contrario, la puntuación no debería superar el 2.

### III. VALIDEZ

Se entiende por validez la capacidad de un instrumento para medir lo que pretende medir. La validez de contenido hace referencia a la adecuación de los ítems y las dimensiones, evaluada por paneles de profesionales y pacientes. La validez de constructo hace referencia a las hipótesis sobre las relaciones con otras medidas y a los patrones de puntuación en grupos conocidos. La validez de criterio hace referencia al grado en el que las puntuaciones del instrumento están asociadas con un criterio de referencia o gold standard.

16. La evidencia disponible sobre validez de contenido para el uso para el que ha sido diseñado el instrumento es suficiente.

- Se evaluarán los siguientes aspectos:
- Evaluación de la validez de contenido en términos de claridad, comprensibilidad, relevancia y redundancia de ítems y escalas.
  - Evaluación de la validez de contenido mediante panel de expertos y/o población diana
  - Si ha habido valoración por solo un tipo de panel, como máximo se podrá puntuar un 2.

Muy de acuerdo  4  3  2  1  Muy en desacuerdo  No hay información

Referencia:

Comentarios:

17. Los métodos empleados para evaluar la validez de constructo y de criterio están suficientemente descritos y son adecuados.

- Se evaluarán los siguientes aspectos:
- Descripción del cálculo del tamaño y obtención de la muestra.
  - Características de la muestra (socio-demográficas, clínicas si proviene de una población de pacientes, etc.)
  - Las condiciones de administración (cómo y dónde fue administrado el instrumento)
  - Estadística descriptiva de los resultados obtenidos con el instrumento (medias, desviación estándar, efectos suelo y techo).

Muy de acuerdo  4  3  2  1  Muy en desacuerdo  No hay información

Referencia:

Comentarios:

#### RECUERDE

Se valorará si se aporta suficiente información, los métodos son adecuados y los resultados obtenidos indican buen funcionamiento del instrumento. Se puntuará como 4 o 3 si se cumplen todos o la mayoría de los criterios indicados. En caso contrario, la puntuación no debería superar el 2.

H

18. La composición de la muestra utilizada para examinar la validez de constructo y de criterio está suficientemente descrita y es adecuada.

Se han considerado con suficiente detalle los factores que podrían afectar a la validez, como el género, la edad, la etnia o la lengua.

Muy de acuerdo  4  3  2  1  Muy en desacuerdo  No hay información

Referencia:

Comentarios:

19. Las hipótesis previas relativas a la validez de constructo están suficientemente descritas y los resultados son consistentes con las mismas.

- Se evaluarán los siguientes aspectos:
- Las hipótesis previas explícitas sobre las relaciones con otras medidas y sobre el patrón de puntuaciones en grupos conocidos. Si no se aportan hipótesis previas, como máximo se podrá puntuar con un 2
  - La consistencia de los resultados sobre la correlación con otras medidas con las hipótesis previas.
  - La consistencia de los resultados sobre el patrón de puntuaciones en grupos conocidos con las hipótesis previas.

Muy de acuerdo  4  3  2  1  Muy en desacuerdo  No hay información

Referencia:

Comentarios:

20. La elección de las medidas de referencia o gold standard para la validez de criterio está suficientemente descrita y razonada.

- Siempre que se evalúen versiones cortas e instrumentos de cribado, deberá aportarse información sobre resultados obtenidos en comparación con las medidas de criterio (sensibilidad, especificidad y curvas ROC). En los otros casos, puede aceptarse que no se evalúe la validez de criterio cuando no exista un instrumento de referencia ampliamente aceptado, la que ocurre con la mayoría de cuestionarios de calidad de vida relacionada con la salud.

Muy de acuerdo  4  3  2  1  Muy en desacuerdo  No aplicable  No hay información

Referencia:

Comentarios:

#### RECUERDE

Se valorará si se aporta suficiente información, los métodos son adecuados y los resultados obtenidos indican buen funcionamiento del instrumento. Se puntuará como 4 o 3 si se cumplen todos o la mayoría de los criterios indicados. En caso contrario, la puntuación no debería superar el 2.

21. La evaluación de la validez del instrumento en cada población de interés está suficientemente descrita.

Si existen razones para pensar que la validez puede variar en determinados grupos o poblaciones (por ejemplo diferentes patologías o grupos culturales) se aportan datos de validez para cada uno de ellos. Si no existen evidencias que sugieran posibles diferencias de validez en las distintas poblaciones se puntuará como no aplicable.

Muy de acuerdo  4  3  2  1  Muy en desacuerdo  No aplicable  No hay información

Referencia:

Comentarios:

#### IV. SENSIBILIDAD AL CAMBIO

Habilidad de un instrumento para detectar cambios en el tiempo

22. Los métodos empleados para evaluar sensibilidad al cambio están suficientemente descritos y son adecuados.

Se evaluarán los siguientes aspectos:

- La descripción del cálculo del tamaño y los métodos de obtención de la muestra.
- Los caracteres de la muestra (socio-demográficos, clínicos si proviene de una población de pacientes, etc.)
- Las condiciones de administración (cómo y dónde fue administrado el instrumento), el intervalo de tiempo y el tipo de intervención (si la hay) entre administraciones.
- Los estadísticos descriptivos de los resultados obtenidos con el instrumento.

Muy de acuerdo  4  3  2  1  Muy en desacuerdo  No hay información

Referencia:

Comentarios:

23. La estimación de la magnitud del cambio está suficientemente descrita y los resultados son

RECUERDE  
Se valorará si se aporta suficiente información, los métodos son adecuados y los resultados obtenidos indican buen funcionamiento del instrumento. Se puntuará como 4 o 3 si se cumplen todos o la mayoría de los criterios indicados. En caso contrario, la puntuación no debería superar el 2.

adecuados.  
Se evaluarán los siguientes aspectos:

- Se proporciona la magnitud del cambio de las puntuaciones del instrumento estimada en estudios longitudinales.
- Se proporcionan medidas ajustadas o estandarizadas del cambio entre antes y después, como por ej.: tamaño del efecto, medias estandarizadas del cambio, EEM (Error estándar de la medida)
- Se indica la fórmula utilizada para calcular los estimadores estadísticos del cambio.

Muy de acuerdo  4  3  2  1  Muy en desacuerdo  No hay información

Referencia:

Comentarios:

24. La magnitud del cambio del grupo que se espera que cambie se ha comparado con la de un grupo estable en estudios longitudinales.

Se evaluarán los siguientes aspectos:

- Se aportan datos longitudinales de los grupos que han presentado cambios.
- Se aportan datos longitudinales de los grupos estables.
- Se evalúa la diferencia entre la magnitud del cambio en ambos grupos.

Como max. se puntuará 3 cuando se comparen grupos provenientes de diferentes estudios.

Muy de acuerdo  4  3  2  1  Muy en desacuerdo  No hay información

Referencia:

Comentarios:

RECUERDE  
Se valorará si se aporta suficiente información, los métodos son adecuados y los resultados obtenidos indican buen funcionamiento del instrumento. Se puntuará como 4 o 3 si se cumplen todos o la mayoría de los criterios indicados. En caso contrario, la puntuación no debería superar el 2.



#### V. INTERPRETABILIDAD

La facilidad para asignar un significado comprensible a las puntuaciones del instrumento. Los métodos más utilizados son:

- "puntuaciones de referencia": procedentes de población general o de otras poblaciones;
- "anclajes" permiten comparar diferencias o cambios en las puntuaciones con otras características (ej., estado clínico, presencia o ausencia de una enfermedad, cambio mínimo importante, evento vital, intervención terapéutica, etc.);
- "criterios externos" que permiten estimar la capacidad predictiva de la puntuación para predecir eventos importantes tales como el fallecimiento o la institucionalización.

25. La selección y la evaluación de los criterios para la interpretación de las puntuaciones están suficientemente descritas y razonadas.

Se evaluarán los siguientes aspectos:

- La justificación de la selección de los criterios externos, los anclajes y las poblaciones de comparación.
- Los métodos de selección de la muestra y cálculo del tamaño muestral.
- Las características de la muestra.
- Las condiciones de administración.

Muy de acuerdo  4  3  2  1  Muy en desacuerdo  No hay información

Referencia:

Comentarios:

#### RECUERDE

Se valorará si se aporta suficiente información, los métodos son adecuados y los resultados obtenidos indican buen funcionamiento del instrumento. Se puntuará como 4 o 3 si se cumplen todos o la mayoría de los criterios indicados. En caso contrario, la puntuación no debería superar el 2.



26. Las estrategias para facilitar la interpretación están suficientemente descritas y son adecuadas.

Se evaluarán los siguientes aspectos:

- Se proporcionan puntuaciones de referencia procedentes de muestras representativas de población general o de otras poblaciones para comparación.
- Se indican "anclajes" que permiten la interpretación de las puntuaciones en función de otras características de los individuos (p.ej., estado clínico, presencia o ausencia de una enfermedad, etc.).
- Se indican "anclajes" que permitan la interpretación de cambios longitudinales (ej., definición del cambio mínimo importante, diferencias asociadas a una determinada intervención o evento vital).
- Se indica la capacidad de predicción de eventos relevantes, como la muerte o la institucionalización.

Muy de acuerdo  4  3  2  1  Muy en desacuerdo  No hay información

Referencia:

Comentarios:

27. La forma en que se deberían presentar los resultados del instrumento está suficientemente descrita.

Se evaluarán los siguientes aspectos:

- El rango teórico y el rango observado.
- Las características de la población de referencia en que se basa para el cálculo de las puntuaciones estandarizadas.

Muy de acuerdo  4  3  2  1  Muy en desacuerdo  No hay información

Referencia:

Comentarios:

#### RECUERDE

Se valorará si se aporta suficiente información, los métodos son adecuados y los resultados obtenidos indican buen funcionamiento del instrumento. Se puntuará como 4 o 3 si se cumplen todos o la mayoría de los criterios indicados. En caso contrario, la puntuación no debería superar el 2.

## VII. FACILIDAD DE USO DEL INSTRUMENTO

Se deberá evaluar de forma general los recursos necesarios para administrar el instrumento (tiempo, esfuerzo y otros). No se han descrito estándares para estas características.

### VI.1. Carga para el/la entrevistado/a.

28. Las habilidades y el tiempo necesarios para completar el instrumento están suficientemente descritos y son aceptables.

Se evaluarán los siguientes aspectos:

- El promedio y rango de tiempo que se necesita para que la persona entrevistada complete el instrumento (de forma auto-administrada, en una entrevista cara a cara, por teléfono o por medio de un ordenador).
- La capacidad de lectura y de comprensión necesarias.
- La necesidad de consultar documentos de salud con el entrevistado (uso de medicamentos, pruebas diagnósticas y otros).
- La adecuación del diseño del instrumento a la población a la cual se dirige (tamaño de letra, opciones de respuesta, saltos, etc.)

Muy de acuerdo  4  3  2  1 Muy en desacuerdo  No hay información

Referencia:

Comentarios:

29. La evaluación de la aceptabilidad y de la capacidad del instrumento para no crear tensión emocional o física indebida están suficientemente descritos y son aceptables.

Se evaluarán los siguientes aspectos:

- Proporción de rechazos (para todo el cuestionario y para los ítems) y las razones.
- Proporción de datos perdidos (para todo el cuestionario y para los ítems) y las razones.
- Modo de administración: (por correo postal, de forma auto-administrada, en una entrevista cara a cara, por teléfono o asistida por ordenador)

Muy de acuerdo  4  3  2  1 Muy en desacuerdo  No hay información

Referencia:

Comentarios:

RECUERDE

Se valorará si se aporta suficiente información, los métodos son adecuados y los resultados obtenidos indican buen funcionamiento del instrumento. Se puntuará como 4 o 3 si se cumplen todos o la mayoría de los criterios indicados. En caso contrario, la puntuación no debería superar el 2.

30. Se indican las circunstancias en las que el instrumento no es apropiado para un/a entrevistado/a.

Se evaluarán los siguientes aspectos:

- Nivel cognitivo mínimo para su aplicación.
- Nivel de lectura en población infantil.
- Rango de edad adecuado para su aplicación.
- Ambito adecuado de aplicación (domicilio, hospitalización e institucionalización).

Muy de acuerdo  4  3  2  1 Muy en desacuerdo  No hay información

Referencia:

Comentarios:

### VI.2. Carga de administración y puntuación del cuestionario.

31. Se detallan los recursos necesarios para la administración del instrumento.

Se deberá evaluar si se considera la necesidad de los siguientes aspectos: manual del entrevistador, material de ayuda (tarjetas con las opciones de respuesta, calendario, fotografías, etc.), necesidad de consultar documentos de salud del entrevistado (listado de medicamentos, pruebas diagnósticas y otros), ordenadores, teléfonos, programas informáticos especiales, para la administración o para el cálculo de las puntuaciones.

Muy de acuerdo  4  3  2  1 Muy en desacuerdo  No hay información

Referencia:

Comentarios:

32. Se especifica el tiempo que necesita un entrevistador/a entrenado/a para administrar el instrumento y es aceptable

Se evaluará el promedio y rango de tiempo que se necesita para que un entrevistador entrenado administre el instrumento.  
En instrumentos autoadministrados se puntuará como no aplicable.

RECUERDE

Se valorará si se aporta suficiente información, los métodos son adecuados y los resultados obtenidos indican buen funcionamiento del instrumento. Se puntuará como 4 o 3 si se cumplen todos o la mayoría de los criterios indicados. En caso contrario, la puntuación no debería superar el 2.

VII. FORMAS ALTERNATIVAS DE ADMINISTRACIÓN

Muy de acuerdo  4  3  2  1  Muy en desacuerdo  No aplicable  No hay información

Referencia:

Comentarios:

33. La formación y experiencia necesarias para la administración del instrumento están suficientemente descritas y son aceptables.

Se evaluarán los siguientes aspectos:

- Tiempo de entrenamiento requerido.
- Nivel académico o de especialización requerido.

En instrumentos autoadministrados se puntuará como no aplicable.

Muy de acuerdo  4  3  2  1  Muy en desacuerdo  No aplicable  No hay información

Referencia:

Comentarios:

34. El método para el cálculo de las puntuaciones está suficientemente descrito y el nivel de carga asociado es aceptable.

Se evaluarán los siguientes aspectos:

- La disponibilidad de un manual de puntuación.
- La necesidad de programas informáticos para el cálculo de la puntuación.
- La necesidad de normas de referencia u otros criterios externos para el cálculo de la puntuación.

Muy de acuerdo  4  3  2  1  Muy en desacuerdo  No aplicable  No hay información

Referencia:

Comentarios:

**RECUERDE**

Se valorará si se aporta suficiente información, los métodos son adecuados y los resultados obtenidos indican buen funcionamiento del instrumento. Se puntuará como 4 o 3 si se cumplen todos o la mayoría de los criterios indicados. En caso contrario, la puntuación no debería superar el 2.

Se consideran formas alternativas de administración cualquiera de las formas para las que no ha sido originalmente diseñado el instrumento: autoadministrado, por entrevistador, entrevista asistida por ordenador u otras formas de administración. También se consideran formas alternativas aquellos instrumentos originalmente diseñados para ser contestados por informadores directos cuando son administrados a informadores/as indirectos o proxies (padres, esposas/os, profesionales, cuidadores etc.).

ATENCIÓN: es posible que el instrumento que está evaluando NO tenga formas de administración alternativas. Indíquelo marcando la siguiente opción:  No aplicable

35. Las características métricas y de uso de cada forma alternativa de administración están suficientemente descritas y son adecuadas.

Se debe evaluar la información disponible para cada forma alternativa de administración relativa a fiabilidad, validez, sensibilidad al cambio, interpretabilidad y facilidad de uso.

Muy de acuerdo  4  3  2  1  Muy en desacuerdo  No aplicable  No hay información

Referencia:

Comentarios:

36. La comparabilidad de las formas alternativas de administración con la original está suficientemente descrita y los resultados son adecuados.

Se evaluarán los siguientes aspectos:

- Comparación de las características métricas de las formas alternativas con la original.
- Valoración de la equivalencia de las puntuaciones obtenidas con las formas alternativas de administración.

Muy de acuerdo  4  3  2  1  Muy en desacuerdo  No aplicable  No hay información

Referencia:

Comentarios:

VIII. PROCESO DE ADAPTACIÓN CULTURAL Y LINGÜÍSTICA

**IMPORTANTE:**

Esta sección se evaluará sólo en caso de instrumentos adaptados.

La adaptación cultural del instrumento implica la evaluación de la equivalencia lingüística, conceptual y de las propiedades de medida.

37. Los métodos empleados para alcanzar y evaluar la equivalencia lingüística con la versión original están suficientemente descritos y son adecuados

Se evaluarán los siguientes aspectos:

- o Se han realizado al menos dos traducciones independientes de la lengua de origen a la lengua que se quiere adaptar.
- o Se ha realizado por lo menos una retro-traducción a la lengua de origen, preferiblemente más de una.
- o Las versiones traducidas han sido revisadas por paneles de profesionales y de pacientes (de población a la que va dirigido).
- o Se ha de evaluar si se han realizado estudios de campo que proporcionan evidencia de la aceptabilidad y comprensión del instrumento.

Muy de acuerdo      1 Muy en desacuerdo

Referencia:

Comentarios:

38. Los métodos empleados para alcanzar y evaluar la equivalencia conceptual de los ítems están suficientemente descritos y son adecuados

Se evaluarán los siguientes aspectos:

- Se ha evaluado la validez de contenido (similitudes en relevancia y significado de los mismos conceptos medidos) en cada grupo cultural o lingüístico a los que se aplicará, mediante paneles de profesionales y de pacientes (o población a la que va dirigido).
- Se ha estudiado el eventual funcionamiento diferencial de los ítems (algunos ítems podrían funcionar de manera diferente en el original y en la versión adaptada del instrumento), p. ej. mediante aproximaciones de teoría de respuesta al ítem, análisis factorial exploratorio o confirmatorio u otros.

Muy de acuerdo      1 Muy en desacuerdo

Referencia:

Comentarios:

39. Las diferencias relevantes entre el original y la versión adaptada están suficientemente descritas y se han resuelto de manera satisfactoria.

Se evaluarán los siguientes aspectos:

- Se han identificado las diferencias más relevantes entre el original y la versión adaptada.
- Se razonan estas diferencias.
- Se han conlucido las incongruencias de manera satisfactoria.

Muy de acuerdo      1 Muy en desacuerdo

Referencia:

Comentarios:

MUCHAS GRACIAS POR CONTESTAR ESTE CUESTIONARIO

## Anexo II:

### Artículos incluidos en el estudio de evaluación de CMS (capítulo 8)

#### 1. Patología Subacromial: 37 artículos

- 1 Urvoy P, Boileau G, Berger M, Vanvelcenaher J, Schmidt D, Herlant M, et al. [Correlation and validation of different methods of evaluation of results after surgery of the rotator cuff. Plea for a standardized method]. *Rev Chir Orthop Reparatrice Appar Mot* 1991;77(3):171-8.
- 2 Gazielly DF, Gleyze P, Montagnon C. Functional and anatomical results after rotator cuff repair. *Clin Orthop Relat Res* 1994 Jul;(304):43-53.
- 3 Gazielly DF, Gleyze P, Montagnon C, Bruyere G, Prallet B. [Functional and anatomical results after surgical treatment of ruptures of the rotator cuff. 1: Preoperative functional and anatomical evaluation of ruptures of the rotator cuff]. *Rev Chir Orthop Reparatrice Appar Mot* 1995;81(1):8-16.
- 4 Boussagol B, Pelissier J, Herisson C, Simon L. [Clinical assessment of rotator cuff lesions: Comparative analysis of four scales]. *Ann Readapt Med Phys* 1996;39(2):79-88.
- 5 T'Jonck L, Lysens R, De SL, Bellemans J, Stoffelen D, Tirez B, et al. Open versus arthroscopic subacromial decompression: analysis of one-year results. *Physiother Res Int* 1997;2(2):46-61.
- 6 Gartsman GM, Brinker MR, Khan M. Early effectiveness of arthroscopic repair for full-thickness tears of the rotator cuff: an outcome analysis. *J Bone Joint Surg Am* 1998 Jan;80(1):33-40.
- 7 O'Connor DA, Chipchase LS, Tomlinson J, Krishnan J. Arthroscopic subacromial decompression: responsiveness of disease-specific and health-related quality of life outcome measures. *Arthroscopy* 1999 Nov;15(8):836-40.
- 8 Skutek M, Fremerey RW, Zeichen J, Bosch U. Outcome analysis following open rotator cuff repair. Early effectiveness validated using four different shoulder assessment scales. *Arch Orthop Trauma Surg* 2000;120(7-8):432-6.
- 9 Torrens C, Orient F, Vila F, Escalada F, Marín M. Análisis comparativo entre la escala de Constant y el cuestionario de salud SF-36 en pacientes con patología subacromial. *Revista de Ortopedia y Traumatología* 2000;44:447-51.
- 10 Dawson J, Hill G, Fitzpatrick R, Carr A. The benefits of using patient-based methods of assessment. Medium-term results of an observational study of shoulder surgery. *J Bone Joint Surg Br* 2001 Aug;83(6):877-82.
- 11 Skutek M, Zeichen J, Fremerey RW, Bosch U. [Outcome analysis after open reconstruction of rotator cuff ruptures. A comparative assessment of recent evaluation procedures]. *Unfallchirurg* 2001 Jun;104(6):480-7.
- 12 Dawson J, Hill G, Fitzpatrick R, Carr A. Comparison of clinical and patient-based measures to assess medium-term outcomes following shoulder surgery for disorders of the rotator cuff. *Arthritis Rheum* 2002 Oct 15;47(5):513-9.
- 13 Kirkley A, Alvarez C, Griffin S. The development and evaluation of a disease-specific quality-of-life questionnaire for disorders of the rotator cuff: The Western Ontario Rotator Cuff Index. *Clin J Sport Med* 2003 Mar;13(2):84-92.
- 14 Huber W, Hofstaetter JG, Hanslik-Schnabel B, Posch M, Wurnig C. The German version of the Oxford Shoulder Score--cross-cultural adaptation and validation. *Arch Orthop Trauma Surg* 2004 Oct;124(8):531-6.
- 15 Porcellini G, Paladini P, Campi F, Paganelli M. Arthroscopic treatment of calcifying tendinitis of the shoulder: clinical and ultrasonographic follow-up findings at two to five years. *J Shoulder Elbow Surg* 2004 Sep;13(5):503-8.
- 16 Romeo AA, Mazzocca A, Hang DW, Shott S, Bach BR, Jr. Shoulder scoring scales for the evaluation of rotator cuff repair. *Clin Orthop Relat Res* 2004 Oct;(427):107-14.
- 17 Holtby R, Razmjou H. Measurement properties of the Western Ontario rotator cuff outcome measure: a preliminary report. *J Shoulder Elbow Surg* 2005 Sep;14(5):506-10.
- 18 Huber W, Hofstaetter JG, Hanslik-Schnabel B, Posch M, Wurnig C. [Translation and psychometric testing of the Rotator Cuff Quality-of-Life Measure (RC-QOL) for use in German-speaking regions]. *Z Rheumatol* 2005 Apr;64(3):188-97.

- 19 Huber W, Hofstaetter JG, Hanslik-Schnabel B, Posch M, Wurnig C. [Translation and psychometric testing of the Western Ontario Rotator Cuff Index (WORC) for use in Germany]. *Z Orthop Ihre Grenzgeb* 2005 Jul;143(4):453-60.
- 20 Wollmerstedt N, Bohm DT, Kirschner S, Kohler M, Konig A. [Evaluation of a simple test for shoulder function in patients with surgically treated rotator cuff injuries]. *Z Orthop Ihre Grenzgeb* 2005 Jul;143(4):468-74.
- 21 El O, Bircan C, Gulbahar S, Demiral Y, Sahin E, Baydar M, et al. The reliability and validity of the Turkish version of the Western Ontario Rotator Cuff Index. *Rheumatol Int* 2006 Oct;26(12):1101-8.
- 22 Livain T, Pichon H, Vermeulen J, Vaillant J, Saragaglia D, Poisson MF, et al. [Intra- and interobserver reproducibility of the French version of the Constant-Murley shoulder assessment during rehabilitation after rotator cuff surgery]. *Rev Chir Orthop Reparatrice Appar Mot* 2007 Apr;93(2):142-9.
- 23 Ghroubi S, Chaari M, Elleuch H, Guermazi M, Baklouti S, Elleuch MH. [Functional and quality of life outcome of non-operated rotator cuff tears]. *Ann Readapt Med Phys* 2008 Dec;51(9):714-21.
- 24 Gialanella B. [Disability and occupational therapy in patient with tear of the rotator cuff]. *G Ital Med Lav Ergon* 2008 Jan;30(1):78-82.
- 25 Razmjou H, Bean A, Macdermid JC, van O, V, Travers N, Holtby R. Convergent validity of the constant-murley outcome measure in patients with rotator cuff disease. *Physiother Can* 2008;60(1):72-9.
- 26 Tae SK, Rhee YG, Park TS, Lee KW, Park JY, Choi CH, et al. The development and validation of an appraisal method for rotator cuff disorders: the Korean Shoulder Scoring System. *J Shoulder Elbow Surg* 2009 Sep;18(5):689-96.
- 27 Goutallier D, Postel JM, Chevalier X, Beaudreuil J, Zilber S. Intermediate term functional outcome prediction following full thickness rotator cuff tear reparative or not reparative surgery. *Orthop Traumatol Surg Res* 2010 Nov;96(7):727-33.
- 28 Longo UG, Vasta S, Maffulli N, Denaro V. Scoring systems for the functional assessment of patients with rotator cuff pathology. *Sports Med Arthrosc* 2011 Sep;19(3):310-20.
- 29 Wessel RN, Lim TE, van MH, de Bie RA. Validation of the Western Ontario Rotator Cuff index in patients with arthroscopic rotator cuff repair: a study protocol. *BMC Musculoskelet Disord* 2011;12:64.
- 30 Chung SW, Park JS, Kim SH, Shin SH, Oh JH. Quality of life after arthroscopic rotator cuff repair: evaluation using SF-36 and an analysis of affecting clinical factors. *Am J Sports Med* 2012 Mar;40(3):631-9.
- 31 de Witte PB, Henseler JF, Nagels J, Vliet Vlieland TP, Nelissen RG. The Western Ontario rotator cuff index in rotator cuff disease patients: a comprehensive reliability and responsiveness validation study. *Am J Sports Med* 2012 Jul;40(7):1611-9.
- 32 Razmjou H, Stratford P, Holtby R. A shortened version of the Western ontario rotator cuff disability index: development and measurement properties. *Physiother Can* 2012;64(2):135-44.
- 33 Karas V, Hussey K, Romeo AR, Verma N, Cole BJ, Mather RC, III. Comparison of subjective and objective outcomes after rotator cuff repair. *Arthroscopy* 2013 Nov;29(11):1755-61.
- 34 Kukkonen J, Kauko T, Vahlberg T, Joukainen A, Aarimaa V. Investigating minimal clinically important difference for Constant score in patients undergoing rotator cuff surgery. *J Shoulder Elbow Surg* 2013 Dec;22(12):1650-5.
- 35 Wessel RN, Wolterbeek N, Fermont AJ, van MH, Sonneveld H, Griffin S, et al. The conceptually equivalent Dutch version of the Western Ontario Rotator Cuff Index (WORC)(c). *BMC Musculoskelet Disord* 2013;14:362.
- 36 Holmgren T, Oberg B, Adolfsson L, Bjornsson HH, Johansson K. Minimal important changes in the Constant-Murley score in patients with subacromial pain. *J Shoulder Elbow Surg* 2014 Aug;23(8):1083-90.
- 37 Moeller AD, Thorsen RR, Torabi TP, Bjoerkman AS, Christensen EH, Maribo T, et al. The Danish version of the modified Constant-Murley shoulder score: reliability, agreement, and construct validity. *J Orthop Sports Phys Ther* 2014 May;44(5):336-40.

***Articulos adicionales que pertenecen a otro grupo (por incluir datos relevantes)***

- Conboy VB, Morris RW, Kiss J, Carr AJ. An evaluation of the Constant-Murley shoulder assessment. *J Bone Joint Surg Br* 1996; 78(2):229-232.
- Gilbert MK, Gerber C. Comparison of the subjective shoulder value and the Constant score. *J Shoulder Elbow Surg* 2007; 16(6):717-721.

## **2. Fracturas: 7 artículos**

- 1 Tingart M, Bathis H, Lefering R, Bouillon B, Tiling T. [Constant Score and Neer Score. A comparison of score results and subjective patient satisfaction]. *Unfallchirurg* 2001 Nov;104(11):1048-54.
- 2 Baker P, Nanda R, Goodchild L, Finn P, Rangan A. A comparison of the Constant and Oxford shoulder scores in patients with conservatively treated proximal humeral fractures. *J Shoulder Elbow Surg* 2008 Jan;17(1):37-41.
- 3 Poeze M, Lenssen AF, Van Empel JM, Verbruggen JP. Conservative management of proximal humeral fractures: can poor functional outcome be related to standard transscapular radiographic evaluation? *J Shoulder Elbow Surg* 2010 Mar;19(2):273-81.
- 4 Slobogean GP, Slobogean BL. Measuring shoulder injury function: common scales and checklists. *Injury* 2011 Mar;42(3):248-52.
- 5 Urda A, Gonzalez A, Colino A, Lopiz Y, Garcia-Fernandez C, Marco F. Management of displaced surgical neck fractures of the humerus: health related quality of life, functional and radiographic results. *Injury* 2012 Dec;43 Suppl 2:S12-S19.
- 6 Inauen C, Platz A, Meier C, Zingg U, Rufibach K, Spross C, et al. Quality of life after osteosynthesis of fractures of the proximal humerus. *J Orthop Trauma* 2013 Apr;27(4):e74-e80.
- 7 van de Water AT, Shields N, Davidson M, Evans M, Taylor NF. Reliability and validity of shoulder function outcome measures in people with a proximal humeral fracture. *Disabil Rehabil* 2014;36(13):1072-9.

## **3.Artritis: 6 artículos**

- 1 van den Ende CH, Rozing PM, Dijkmans BA, Verhoef JA, Voogt-van der Harst EM, Hazes JM. Assessment of shoulder function in rheumatoid arthritis. *J Rheumatol* 1996 Dec;23(12):2043-8.
- 2 Angst F, Pap G, Mannion AF, Herren DB, Aeschlimann A, Schwyzer HK, et al. Comprehensive assessment of clinical outcome and quality of life after total shoulder arthroplasty: usefulness and validity of subjective outcome measures. *Arthritis Rheum* 2004 Oct 15;51(5):819-28.
- 3 Angst F, Goldhahn J, Drerup S, Aeschlimann A, Schwyzer HK, Simmen BR. Responsiveness of six outcome assessment instruments in total shoulder arthroplasty. *Arthritis Rheum* 2008 Mar 15;59(3):391-8.
- 4 Christie A, Hagen KB, Mowinckel P, Dagfinrud H. Methodological properties of six shoulder disability measures in patients with rheumatic diseases referred for shoulder surgery. *J Shoulder Elbow Surg* 2009 Jan;18(1):89-95.
- 5 Christie A, Dagfinrud H, Garratt AM, Ringen OH, Hagen KB. Identification of shoulder-specific patient acceptable symptom state in patients with rheumatic diseases undergoing shoulder surgery. *J Hand Ther* 2011 Jan;24(1):53-60.
- 6 Rasmussen JV, Jakobsen J, Olsen BS, Brorson S. Translation and validation of the Western Ontario Osteoarthritis of the Shoulder (WOOS) index - the Danish version. *Patient Relat Outcome Meas* 2013;4:49-54.

## **Artículos adicionales que pertenecen a otro grupo (por incluir datos relevantes)**

- Conboy VB, Morris RW, Kiss J, Carr AJ. An evaluation of the Constant-Murley shoulder assessment. *J Bone Joint Surg Br* 1996 Mar;78(2):229-32.
- Gilbert MK, Gerber C. Comparison of the subjective shoulder value and the Constant score. *J Shoulder Elbow Surg* 2007 Nov;16(6):717-21.

## **4.Inestabilidad: 5 artículos**

- 1 Kirkley A, Griffin S, McLintock H, Ng L. The development and evaluation of a disease-specific quality of life measurement tool for shoulder instability. The Western Ontario Shoulder Instability Index (WOSI). *Am J Sports Med* 1998 Nov;26(6):764-72.
- 2 Dawson J, Fitzpatrick R, Carr A. The assessment of shoulder instability. The development and validation of a questionnaire. *J Bone Joint Surg Br* 1999 May;81(3):420-6.
- 3 Plancher KD, Lipnick SL. Analysis of evidence-based medicine for shoulder instability. *Arthroscopy* 2009 Aug;25(8):897-908.
- 4 Hofstaetter JG, Hanslik-Schnabel B, Hofstaetter SG, Wurnig C, Huber W. Cross-cultural adaptation and validation of the German version of the Western Ontario Shoulder Instability index. *Arch Orthop Trauma Surg* 2010 Jun;130(6):787-96.

- 5 Kemp KA, Sheps DM, Beaupre LA, Styles-Tripp F, Luciak-Corea C, Balyk R. An evaluation of the responsiveness and discriminant validity of shoulder questionnaires among patients receiving surgical correction of shoulder instability. *ScientificWorldJournal* 2012;2012:410125.

***Articulos adicionales que pertenecen a otro grupo (por incluir datos relevantes)***

- Gilbert MK, Gerber C. Comparison of the subjective shoulder value and the Constant score. *J Shoulder Elbow Surg* 2007 Nov;16(6):717-21.

**5. Hombro congelado: 1 articulo**

- 1 Othman A, Taylor G. Is the constant score reliable in assessing patients with frozen shoulder? 60 shoulders scored 3 years after manipulation under anaesthesia. *Acta Orthop Scand* 2004 Feb;75(1):114-6.

**6. Varias Patologías: 29 articulos**

- 1 Conboy VB, Morris RW, Kiss J, Carr AJ. An evaluation of the Constant-Murley shoulder assessment. *J Bone Joint Surg Br* 1996 Mar;78(2):229-32.
- 2 Dawson J, Fitzpatrick R, Carr A. Questionnaire on the perceptions of patients about shoulder surgery. *J Bone Joint Surg Br* 1996 Jul;78(4):593-600.
- 3 Kohn D, Geyer M. The subjective shoulder rating system. *Arch Orthop Trauma Surg* 1997;116(6-7):324-8.
- 4 Bankes MJ, Crossman JE, Emery RJ. A standard method of shoulder strength measurement for the Constant score with a spring balance. *J Shoulder Elbow Surg* 1998 Mar;7(2):116-21.
- 5 Taylor RJ, Chepeha JC, Teknos TN, Bradford CR, Sharma PK, Terrell JE, et al. Development and validation of the neck dissection impairment index: a quality of life measure. *Arch Otolaryngol Head Neck Surg* 2002 Jan;128(1):44-9.
- 6 Cook KF, Roddey TS, Olson SL, Gartsman GM, Valenzuela FF, Hanten WP. Reliability by surgical status of self-reported outcomes in patients who have shoulder pathologies. *J Orthop Sports Phys Ther* 2002 Jul;32(7):336-46.
- 7 Boehm D, Wollmerstedt N, Doesch M, Handwerker M, Mehling E, Gohlke F. [Development of a questionnaire based on the Constant-Murley-Score for self-evaluation of shoulder function by patients]. *Unfallchirurg* 2004 May;107(5):397-402.
- 8 Placzek JD, Lukens SC, Badalanmenti S, Roubal PJ, Freeman DC, Walleman KM, et al. Shoulder outcome measures: a comparison of 6 functional tests. *Am J Sports Med* 2004 Jul;32(5):1270-7.
- 9 Fialka C, Oberleitner G, Stampfl P, Brannath W, Hexel M, Vecsei V. Modification of the Constant-Murley shoulder score-introduction of the individual relative Constant score Individual shoulder assessment. *Injury* 2005 Oct;36(10):1159-65.
- 10 Leggin BG, Michener LA, Shaffer MA, Brenneman SK, Iannotti JP, Williams GR, Jr. The Penn shoulder score: reliability and validity. *J Orthop Sports Phys Ther* 2006 Mar;36(3):138-51.
- 11 Gilbert MK, Gerber C. Comparison of the subjective shoulder value and the Constant score. *J Shoulder Elbow Surg* 2007 Nov;16(6):717-21.
- 12 Fayad F, Lefevre-Colau MM, Mace Y, Fermanian J, Mayoux-Benhamou A, Roren A, et al. Validation of the French version of the Disability of the Arm, Shoulder and Hand questionnaire (F-DASH). *Joint Bone Spine* 2008 Mar;75(2):195-200.
- 13 Rocourt MH, Radlinger L, Kalberer F, Sanavi S, Schmid NS, Leunig M, et al. Evaluation of intratester and intertester reliability of the Constant-Murley shoulder assessment. *J Shoulder Elbow Surg* 2008 Mar;17(2):364-9.
- 14 Ozsahin M, Akgun K, Aktas I, Kurtais Y. Adaptation of the Shoulder Disability Questionnaire to the Turkish population, its reliability and validity. *Int J Rehabil Res* 2008 Sep;31(3):241-5.
- 15 Fayad F, Lefevre-Colau MM, Gautheron V, Mace Y, Fermanian J, Mayoux-Benhamou A, et al. Reliability, validity and responsiveness of the French version of the questionnaire Quick Disability of the Arm, Shoulder and Hand in shoulder disorders. *Man Ther* 2009 Apr;14(2):206-12.
- 16 Oh JH, Jo KH, Kim WS, Gong HS, Han SG, Kim YH. Comparative evaluation of the measurement properties of various shoulder outcome instruments. *Am J Sports Med* 2009 Jun;37(6):1161-8.
- 17 Clement ND, Fuller M, Colling RC, Stirrat AN. Assessment of shoulder function using the coronal plane angle. *Int J Shoulder Surg* 2009 Oct;3(4):90-3.
- 18 Berendes T, Pilot P, Willems J, Verburg H, te SR. Validation of the Dutch version of the Oxford Shoulder Score. *J Shoulder Elbow Surg* 2010 Sep;19(6):829-36.



- 19 Murena L, Vulcano E, D'Angelo F, Monti M, Cherubino P. Italian cross-cultural adaptation and validation of the Oxford Shoulder Score. *J Shoulder Elbow Surg* 2010 Apr;19(3):335-41.
- 20 Dupeyron A, Gelis A, Sablayrolles P, Bousquet PJ, Julia M, Herisson C, et al. Heterogeneous assessment of shoulder disorders: validation of the Standardized Index of Shoulder Function. *J Rehabil Med* 2010 Nov;42(10):967-72.
- 21 Blonna D, Scelsi M, Marini E, Bellato E, Tellini A, Rossi R, et al. Can we improve the reliability of the Constant-Murley score? *J Shoulder Elbow Surg* 2012 Jan;21(1):4-12.
- 22 Frich LH, Noergaard PM, Brorson S. Validation of the Danish version of Oxford Shoulder Score. *Dan Med Bull* 2011 Nov;58(11):A4335.
- 23 Noorani AM, Roberts DJ, Malone AA, Waters TS, Jaggi A, Lambert SM, et al. Validation of the Stanmore percentage of normal shoulder assessment. *Int J Shoulder Surg* 2012 Jan;6(1):9-14.
- 24 Bafus BT, Hughes RE, Miller BS, Carpenter JE. Evaluation of utility in shoulder pathology: Correlating the American Shoulder and Elbow Surgeons and Constant scores to the EuroQoL. *World J Orthop* 2012 Mar 18;3(3):20-4.
- 25 van Kampen DA, van Beers LW, Scholtes VA, Terwee CB, Willems WJ. Validation of the Dutch version of the Simple Shoulder Test. *J Shoulder Elbow Surg* 2012 Jun;21(6):808-14.
- 26 Roh YH, Noh JH, Oh JH, Baek GH, Gong HS. To what degree do shoulder outcome instruments reflect patients' psychologic distress? *Clin Orthop Relat Res* 2012 Dec;470(12):3470-7.
- 27 Schmidutz F, Beirer M, Braunstein V, Bogner V, Wiedemann E, Biberthaler P. The Munich Shoulder Questionnaire (MSQ): development and validation of an effective patient-reported tool for outcome measurement and patient safety in shoulder surgery. *Patient Saf Surg* 2012;6(1):9.
- 28 Ge Y, Chen S, Chen J, Hua Y, Li Y. The development and evaluation of a new shoulder scoring system based on the view of patients and physicians: the Fudan University shoulder score. *Arthroscopy* 2013 Apr;29(4):613-22.
- 29 Levy O, Haddo O, Massoud S, Mullett H, Atoun E. A patient-derived Constant-Murley score is comparable to a clinician-derived score. *Clin Orthop Relat Res* 2014 Jan;472(1):294-303.

### **7. Sujetos Sanos: 9 artículos**

- 1 Brinker MR, Cuomo JS, Popham GJ, O'Connor DP, Barrack RL. An examination of bias in shoulder scoring instruments among healthy collegiate and recreational athletes. *J Shoulder Elbow Surg* 2002; 11(5):463-469.
- 2 Grassi FA, Tajana MS. The normalization of data in the Constant-Murley score for the shoulder. A study conducted on 563 healthy subjects. *Chir Organi Mov* 2003; 88(1):65-73.
- 3 Thomas M, Dieball O, Busse M. [Normal values of the shoulder strength in dependency on age and gender--comparison with the constant, UCLA, ASES scores and SF36 health survey]. *Z Orthop Ihre Grenzgeb* 2003; 141(2):160-170.
- 4 Johansson KM, Adolfsson LE. Intraobserver and interobserver reliability for the strength test in the Constant-Murley shoulder assessment. *J Shoulder Elbow Surg* 2005; 14(3):273-278.
- 5 Katolik LI, Romeo AA, Cole BJ, Verma NN, Hayden JK, Bach BR. Normalization of the Constant score. *J Shoulder Elbow Surg* 2005; 14(3):279-285.
- 6 Yian EH, Ramappa AJ, Arneberg O, Gerber C. The Constant score in normal shoulders. *J Shoulder Elbow Surg* 2005; 14(2):128-133.
- 7 Walton MJ, Walton JC, Honorez LA, Harding VF, Wallace WA. A comparison of methods for shoulder strength assessment and analysis of Constant score change in patients aged over fifty years in the United Kingdom. *J Shoulder Elbow Surg* 2007; 16(3):285-289.
- 8 Tavakkolizadeh A, Ghassemi A, Colegate-Stone T, Latif A, Sinha J. Gender-specific Constant score correction for age. *Knee Surg Sports Traumatol Arthrosc* 2009; 17(5):529-533.
- 9 Hirschmann MT, Wind B, Amsler F, Gross T. Reliability of shoulder abduction strength measure for the Constant-Murley score. *Clin Orthop Relat Res* 2010; 468(6):1565-1571.

### **Artículos adicionales que pertenecen a otro grupo (por incluir datos relevantes)**

- Bankes MJ, Crossman JE, Emery RJ. A standard method of shoulder strength measurement for the Constant score with a spring balance. *J Shoulder Elbow Surg* 1998; 7(2):116-121.
- Fialka C, Oberleitner G, Stampfl P, Brannath W, Hexel M, Vecsei V. Modification of the Constant-Murley shoulder score-introduction of the individual relative Constant score Individual shoulder assessment. *Injury* 2005; 36(10):1159-1165.

**Atributo del concepto y modelo de medida: 2articulos**

- 1 Constant CR, Murley AH. A clinical method of functional assessment of the shoulder. *Clin Orthop Relat Res.* 1987;( 214):160-4.
- 2 Constant CR, Gerber C, Emery RJ, Sjøbjerg JO, Gohlke F, Boileau P. A review of the Constant score: modifications and guidelines for its use. *J Shoulder Elbow Surg.* 2008 Mar-Apr;17 (2):355-61.



