

# *Instrumentos PRO y Niveles de Evidencia en Cirugía de la Mano*

**Tesis Doctoral**

Alfonsa Yolanda Martín Hidalgo



*Universidad de La Laguna, 2015*



**UNIVERSIDAD DE LA LAGUNA**  
**FACULTAD DE CIENCIAS DE LA SALUD**

Memoria presentada por:

Alfonsa Yolanda Martín Hidalgo

Licenciada en Medicina y Cirugía, para optar al título de Doctor  
en Medicina por la Universidad de La Laguna

Fdo. A. Yolanda Martín Hidalgo



© UNIVERSIDAD DE LA LAGUNA





UNIVERSITY DE LA LAGUNA

FACULTY OF MEDICINE

DOCTORAL THESIS

PRO INSTRUMENTS AND LEVEL OF EVIDENCE IN  
HAND SURGERY

A. Yolanda Martín Hidalgo



© UNIVERSIDAD DE LA LAGUNA





Luis Enrique Reboso Morales, Doctor en Medicina por la Universidad de La Laguna y profesor asociado de Traumatología y Cirugía Ortopédica de La Universidad de La Laguna.

**CERTIFICA:**

Que Dña. Alfonsa Yolanda Martín Hidalgo Licenciada en Medicina y Cirugía, ha realizado bajo mi dirección el trabajo titulado “Instrumentos Pro y Niveles de Evidencia en Cirugía de la Mano”

Revisado el presente trabajo y por creer que reúne las condiciones necesarias, estimo que puede ser presentado como Tesis Doctoral.



Roberto Sánchez Rosales Doctor en Medicina y Cirugía. Former Research Fellow of The Kleinert Institute for Hand Surgery, Louisville Ky. Fellowship in Hand and Micro Surgery, University of Louisville Kentucky, USA. Jefe de la Unidad de Cirugía de La Mano y Microcirugía. Tenerife. GECOT. Delegado Internacional de la Sociedad Española de Cirugía de La Mano (SECMA) en la Federación Europea (FESSH) y en la Internacional (IFSSH).

**CERTIFICA:**

Que Dña. Alfonsa Yolanda Martín Hidalgo Licenciada en Medicina y Cirugía, ha realizado bajo mi dirección el trabajo titulado “Instrumentos PRO y Niveles de Evidencia en Cirugía de la Mano”

Revisado el presente trabajo y por creer que reúne las condiciones necesarias, estimo que puede ser presentado como Tesis Doctoral.



Sebastián Rodríguez Hernández, profesor titular de Traumatología y Cirugía Ortopédica de La Universidad de La Laguna.

**CERTIFICA:**

Que Dña. Alfonsa Yolanda Martín Hidalgo Licenciada en Medicina y Cirugía, ha realizado bajo mi dirección el trabajo titulado “Instrumentos PRO y Niveles de Evidencia en Cirugía de la Mano”

Revisado el presente trabajo y por creer que reúne las condiciones necesarias, estimo que puede ser presentado como Tesis Doctoral.



ISAM ATROSHI, Departments of Clinical Sciences Lund University,  
Lund, and Orthopedics Hässleholm-Kristianstad, Hässleholm Hospital,  
Hässleholm. Sweden.

**CERTIFICA:**

Que Dña. Alfonsa Yolanda Martín Hidalgo Licenciada  
en Medicina y Cirugía, ha realizado bajo mi dirección  
el trabajo titulado “Instrumentos PRO y Niveles de  
Evidencia en Cirugía de la Mano”

Revisado el presente trabajo y por creer que reúne las  
condiciones necesarias, estimo que puede ser  
presentado como Tesis Doctoral.



## **AGRADECIMIENTOS**

Este trabajo va dedicado a personas muy especiales, que siempre han estado a mi lado.

Gracias a mis tutores, los Doctores Luis Rebozo y Roberto Sánchez Rosales. Gracias por su paciencia, dedicación, motivación y criterio. Ha sido un orgullo poder contar con su ayuda.

Gracias en especial al Dr. Luis Rebozo, compañero de trabajo, amigo y director de mi Tesis, por su ayuda, orientaciones, persistencia y apoyo, que han sido fundamentales para llevar a cabo este proyecto.

Gracias al Dr Sebastián Rodríguez, de quien tuve el placer de recibir mis primeras enseñanzas durante el inicio de mi vida profesional y del que aprendí sobre todo a cómo ser mejor y a no perder la sonrisa.

Gracias al Dr Atroshi por ser Codirector de mi Tesis Doctoral, es un privilegio.

Gracias a las personas que, de una manera u otra, han sido claves en mi vida profesional, y a mis compañeros de trabajo, en especial al Dr Rayco García.

Gracias a los amigos a los que he robado horas de compañía. Nombrar a todos sería muy extenso y podría cometer algún olvido injusto, pero en especial no puedo dejar de nombrar a Estefanía y María, que me han hecho reír aún cuando era imposible.

Y por encima de todo, y con todo mi amor, gracias a los míos por estar incondicionalmente conmigo, y a los que ya no están. A mi padre, a mi madre, a mi hermano y sobrinos (Adrián y Adara) porque me hacen sentir feliz cada momento que vivo a su lado.



## AGRADECIMIENTOS

## INDICE GENERAL

PUBLICACIONES .....	5
ABREVIATURAS.....	11
1 INTRODUCCIÓN .....	15
1.1 Medicina Basada en la Evidencia .....	15
1.1.1 Niveles de Evidencia .....	19
1.1.1.1 Estudios Descriptivos .....	20
1.1.1.1.1 La Opinión del experto: .....	20
1.1.1.1.2 La observación personal:.....	21
1.1.1.1.3 Series de casos: .....	21
1.1.1.1.4 Estudios de corte o transversales: .....	21
1.1.1.2 Estudios Observacionales .....	22
1.1.1.2.1 Estudios de casos y controles:.....	22
1.1.1.2.2 Estudios de cohortes o longitudinales: .....	23
1.1.1.3 Ensayos Clínicos Controlados Aleatorizados .....	24
1.1.1.4 El Meta-Análisis y La Revisión Sistemática .....	25
1.1.2 Finalidad de la Medicina Basada en la Evidencia .....	26
1.1.2.1 La Colaboración Cochrane .....	27
1.1.3 ¿Cuál es la calidad en las publicaciones de los Ensayos Clínicos Aleatorizados en la Extremidad Superior? .....	29
1.1.4 Y ante estos datos, ¿qué podemos hacer para mejorar nuestros trabajos de investigación clínica en cirugía de la mano? .....	35
1.2 Movimiento Outcomes.....	41
1.3 Instrumentos de Salud (Instrumentos PRO).....	47
1.3.1 Tipos de Instrumentos de Salud .....	49
1.3.1.1 Instrumentos Genéricos .....	49
1.3.1.2 Instrumentos Específicos .....	50
1.3.2 ¿Cómo se desarrolla un Instrumento? .....	51
1.3.3 Instrumentos de Salud Genéricos en la Valoración de Resultados en el Miembro Superior .....	59
1.3.3.1 Cuestionario SF-36 .....	60

1.3.3.2	Cuestionario Euroqol-5D .....	65
1.3.4	Instrumentos de Salud Específicos en la Valoración de Resultados en la Extremidad Superior.....	67
1.3.4.1	Instrumento Michigan Hand Outcomes .....	68
1.3.4.2	Instrumento DASH .....	70
1.3.4.2.1	Versión Española del Instrumento DASH. Adaptación Cultural y Lingüística.....	74
1.3.4.3	Cuestionario QuickDASH .....	78
1.3.4.4	Instrumento Específico del Síndrome del Túnel Carpiano .....	80
1.4	Justificación y Estructura de la Tesis .....	89
1.	Hipótesis Primera .....	93
2.	Hipótesis Segunda .....	93
3.	Hipótesis Tercera.....	94
2	MATERIAL Y MÉTODO.....	95
2.1	Hipótesis Primera: Estudio del Nivel de Evidencia en Cirugía de la Mano .....	95
2.1.1	Población de estudio.....	96
2.1.2	Evaluación del Nivel de Evidencia.....	96
2.1.3	Análisis de la confiabilidad .....	97
2.1.4	Análisis estadístico .....	98
2.2	Hipótesis Segunda: Estudio de las propiedades psicométricas de la versión española del DASH en la valoración de resultados de pacientes operados del Síndrome del Túnel Carpiano .....	99
2.2.1	Población estudio.....	99
2.2.2	Diseño Clínico .....	100
2.2.3	Instrumentos y Medidas.....	100
2.2.4	Análisis de datos.....	101
2.3	Hipótesis tercera: Estudio de las propiedades psicométricas de la versión corta española del Instrumento 6-CTS .....	103
2.3.1	Población estudio.....	103
2.3.2	Diseño Clínico .....	104
2.3.3	Instrumentos y Medidas.....	104
2.3.4	Análisis de Datos .....	106
3	RESULTADOS.....	107
3.1.	Hipótesis Primera: Estudio del Nivel de Evidencia en Cirugía de la Mano .....	107

3.2 Hipótesis Segunda: Estudio de las propiedades psicométricas de la versión española del DASH en la valoración de resultados de pacientes operados del Síndrome del Túnel Carpiano .....	111
3.3 Hipótesis tercera: Estudio de las propiedades psicométricas de la versión corta española del Instrumento 6-CTS .....	115
4 DISCUSIÓN .....	119
4.1 Hipótesis Primera: Estudio del Nivel de Evidencia en Cirugía de la Mano .....	119
4.2 Hipótesis Segunda: Estudio de las propiedades psicométricas de la versión española del DASH en la valoración de resultados de pacientes operados del Síndrome del Túnel Carpiano .....	129
4.3 Hipótesis tercera: Estudio de las propiedades psicométricas de la versión corta española del instrumento 6-CTS .....	137
5 CONCLUSIONES .....	149
6 ANEXOS .....	151
Anexo A: Tabla 1 Niveles de Evidencia Científica según el tipo de estudio de Investigación Clínica Oxford Centre for Evidence-Based Medicine Levels of Evidence (May 2001).....	153
Anexo B: artículo original .....“Level of evidence in hand surgery”	159
Anexo C: artículo original . “The Relative Responsiveness and Construct Validity of the Spanish Version of the DASH Instrument for Outcomes Assessment in Open Carpal Tunnel Release”	175
Anexo D: artículo original .....“Reliability and Construct Validity of the Spanish Version of the 6-item CTS symptoms scale for outcomes assessment in Carpal Tunnel Syndrome”	189
Anexo E: CUESTIONARIO SF-36 .....	203
Anexo F: Versión Española del Instrumento DASH .....	217
Anexo G: QUICKDASH.....	223
Anexo H: Cuestionario de Salud Euroqol-5D.....	225
Anexo I: Versión Española del Instrumento CTS.....	229
Anexo J: Versión Española del Instrumento 6-CTS .....	233
7 BIBLIOGRAFÍA .....	235



## **PUBLICACIONES**

La Tesis Doctoral que aquí se presenta ha dado lugar a las siguientes Publicaciones:

1. "The Relative Responsiveness and Construct Validity of the Spanish Version of the DASH Instrument for Outcomes Assessment in Open Carpal Tunnel Release": publicado en The Journal of Hand Surgery European Vol. 2009;34:72-5
2. "Level of Evidence in Hand Surgery" publicado en la revista Biomedcentral BMC Research Notes 2012, 5:665.
3. "Nivel de evidencia de la RICMA comparada con el JHSE y JHSA": publicado en la revista IberoAmericana de Cirugía de la Mano 2014, 42(2): X-XX

Publicaciones en fase de revisión:

- 1."Reliability and Construct Validity of the Spanish Version of the 6-item CTS symptoms scale for outcomes assessment in Carpal Tunnel Syndrome"

La Tesis Doctoral que aquí se presenta ha dado lugar a las siguientes Comunicaciones:

1. Validez de Construcción de la Versión Española del Instrumento QuickDASH para medir resultados en cirugía de la mano y su valor predictivo en determinar discapacidad de MMSS .XVIII Congreso Nacional de la Sociedad Española de Cirugía de Mano (SECMA 2007).
2. Validez de Construcción del Módulo Específico del cuestionario DASH para medir discapacidad específica de MMSS en Músicos. XIX Congreso Nacional de la Sociedad Española de Cirugía de la Mano (SECMA 2009).
3. Análisis de la fiabilidad (error intra-observador e inter-observador) de la clasificación internacional de la evidencia científica para catalogar la evidencia de las publicaciones en investigación clínica en la revista Ibero-Americana y The Journal Hand Surgery: XX Congreso Nacional de la Sociedad Española de Cirugía de la Mano (SECMA 2011).
4. Fiabilidad y Validez de Construcción de la Versión Española del Instrumento “6-ítems CTS” para valorar Resultados en Síndrome del Túnel Carpiano. XXII Congreso Nacional de la Cirugía de la Mano (SECMA 2015).
5. Reliability and Construct Validity of the 6-item CTS symptoms scale for outcomes assessment in Carpal Tunnel Syndrome. 70th Congreso de la Sociedad Americana de Cirugía de la Mano (ASSH 2015)

6. Reliability and Construct Validity of the 6-item CTS instrument for outcomes assessment in carpal tunnel syndrome. Congreso de La Federación Europea de la Sociedad de Cirugía de la Mano (FESSH 2015)

La tesis Doctoral que aquí se presenta ha dado lugar a las siguientes presentaciones con Póster:

1. “La versión Española de los Instrumentos específicos de Salud DASH (Disability of the Arm, Shoulder and Hand) and CTS (Carpal Tunnel Syndrome) : Adaptación Transcultural y Fiabilidad". I Jornada Científica BiblioPRO Barcelona 2012.
2. “Respuesta Relativa y Validez de construcción de la versión española del Instrumento DASH para valorar resultados en la Liberación del Túnel Carpiano" I Jornada Científica BiblioPRO Barcelona 2012.
3. Validez de Construcción de la Versión Española del Instrumento “Quick DASH” para medir resultados en Cirugía de la mano y su valor predictivo en determinar discapacidad de MMSS". I Jornada Científica BiblioPRO Barcelona 2012.
4. “Validez de Construcción del módulo específico del DASH (Disability of the Arm, Shoulder, and Hand) para medir discapacidad de MMSS en músicos". I Jornada Científica BiblioPRO Barcelona 2012.
5. “Síntomas, Discapacidad de MMSS y Calidad de Vida en el Músico” I Jornada Científica BiblioPRO Barcelona 2012.

6. Reliability and Construct Validity of the 6-item CTS symptoms scale for outcomes assessment in Carpal Tunnel Syndrome. Congreso de la Sociedad Americana de Cirugía de la Mano (ASSH 2015).
7. Responsiveness of the 6-items CTS, Quick DASH and EQ-5D Index Health Instruments for Outcomes Assessment in Carpal Tunnel. XX Congreso de La Federación Europea de la Sociedad de Cirugía de la Mano (FESSH 2015).



## **ABREVIATURAS**

- AAOS: American Academy of Orthopedic Surgeons
- AHRQ: The Agency for Healthcare Research and Quality
- CVRS : Calidad de vida relacionada con la salud
- CONSORT :The Consolidated Standars for Reporting Trials
- CTS: cuestionario Brigham del túnel carpiano
- ECA: Ensayo Clínico Aleatorizado
- ENMG: Electroneuromiograma
- ePRO : Evaluación de Resultados Percibidos
- EVA: escala visual analógica
- DASH: Disabilities of the Arm, Shoulder and Hand
- ICF: International Classification of Functioning, and Disability
- IQLA: International Quality of Life Assessment
- IWH: Institute for Work and Health
- IRYSS: Red de Investigación en Resultados y Servicios Sanitarios
- SPADI: Shoulder Pain and Disability Index
- MBE: Medicina Basada en la Evidencia
- MA: Meta- Análisis
- MHQ: Michigan Hand Questionnaire
- MOS: Medical Outcomes Study
- OSS: Oxford Shoulder Scores
- PEM: Patient Evaluation Measure
- PRWE: Patient Rated Wrist Evaluation

- PRO: Patient Reported Outcomes
- QALY: Quality-Adjusted Life-Years
- ROC: Received Operating Characteristic
- RS: Revisión Sistemática
- SRM: Standarized Response Mean
- STC: Síndrome del Túnel Carpiano
- ES: tamaño del efecto
- WOOS: Western Ontario Osteoarthritis of the Shoulder Index

*"El amor es la alegría de los  
buenos, la reflexión de los  
sabios, y el asombro de los  
incrédulos" Platón.*



# 1

## INTRODUCCIÓN

### 1.1 Medicina Basada en la Evidencia

La Medicina Basada en la Evidencia (MBE) es la utilización meticulosa, explícita y sensata de las mejores evidencias existentes para ayudar a tomar decisiones referentes al tratamiento de los pacientes. La práctica de la MBE implica integrar las experiencias clínicas individuales y las mejores evidencias derivadas de investigaciones sistemáticas (Sackett 1996). Es una de las expresiones de un antiguo movimiento de mejora de la calidad clínica, que se vio potenciada por la demostración de la falta de fundamento de muchas decisiones clínicas y de la enorme variabilidad de la práctica médica. Hoy en día es reconocida en todo el mundo como un progreso a la hora de ofrecer la mejor calidad en la práctica clínica.

Su desarrollo se inició en los años sesenta del siglo XX cuando se empezó a aplicar el diseño de los ensayos clínicos al campo médico. En 1972 Cochrane señaló la falta de fundamento de las decisiones clínicas, Wennberg en 1973 publicó sus trabajos sobre la variabilidad de la práctica clínica, y en 1991 llegó el desarrollo explosivo de la medicina basada en pruebas, primero en un círculo restringido, y después en todo el mundo, por influencia del grupo canadiense de medicina basada en la evidencia (Evidence - Based Medicine Working Group), iniciativa liderada por Sackett. En noviembre de 1992 el "Evidence - Based Medicine Working Group", Grupo de Trabajo en Medicina Basada en la Evidencia de la Universidad de McMaster en Ontario (Canadá), publicaba en la revista JAMA el artículo titulado: "Evidence - Based Medicine: A new approach to teaching the practice of medicine". Se iniciaba la difusión de un nuevo enfoque para la práctica de la medicina. Las metodologías empleadas para determinar la mejor evidencia fueron establecidas por el equipo de McMaster conducido por los médicos David Sackett y Gordon Guyatt en 1990 (Universidad de McMaster, Canadá). Se trataba de crear una medicina científica centrada en el paciente, que resolviera problemas sin crear otros.

La Medicina Basada en la Evidencia, resulta un tema de máximo interés para clínicos, médicos de salud pública, gerentes, planificadores y para el público en general. Pero este entusiasmo se ha mezclado con algunas reacciones en contra. Las críticas van desde que se trata de una "vieja historia", a que es una innovación peligrosa para justificar los recortes de costes y suprimir la libertad clínica. Pero la MBE no es la medicina del recorte de costes. Algunos temen que la MBE sea secuestrada por proveedores y gerentes para recortar los gastos en la atención de la salud. Este no sólo sería un mal uso, sino que sugiere un malentendido fundamental de sus consecuencias económicas.(Sackett 1996)

Los médicos y cirujanos, estamos constantemente introduciendo nuevos productos, nuevas operaciones y nuevos métodos de tratamiento a los pacientes. Estamos bajo la influencia de los nuevos avances tecnológicos que nos eclipsan. Normalmente, lo que usamos como evidencia es la opinión de expertos en nuestro campo, pero ¿cómo sabemos que sus decisiones son correctas?. La práctica de la medicina se ha convertido en el arte de hacer decisiones adecuadas basadas en información inadecuada. Por ejemplo, ¿cuántos de nosotros hemos prescrito condroitín-sulfato para el tratamiento del dolor en la osteoartrosis?. Recientemente, un meta-análisis indica que el beneficio es mínimo o nulo. Y la siguiente pregunta sería, ¿realizará la casa comercial un artículo informando de dichos resultados ?. (Robert M. Szabo 2008)

Es importante mantener en mente lo que Sackett defendió; la evidencia es el uso consciente, explícito y juicioso de las mejores y más actuales evidencias o pruebas en la toma de las decisiones sobre el cuidado de los pacientes. Practicar la MBE significa integrar la competencia clínica individual, con la mayor evidencia clínica externa disponible, a partir de la investigación sistemática. Esta evidencia clínica externa puede informar, pero nunca reemplazar la experiencia clínica individual. Por competencia clínica individual se entiende la habilidad y buen juicio que el clínico adquiere a través de la experiencia y la práctica clínica. La mayor competencia se refleja de muchas maneras, pero sobre todo en un diagnóstico más efectivo y eficiente, y en una identificación más prudente y manejo compasivo de los problemas, derechos y preferencias del paciente individual, al tomar decisiones clínicas acerca de su cuidado. Por mejor evidencia clínica externa disponible, se quiere significar la investigación clínicamente relevante, a menudo desde las ciencias básicas de la medicina, pero sobre todo desde la investigación clínica centrada en el paciente, hacia la exactitud y precisión

de las pruebas diagnósticas (incluido el examen clínico), la potencia de los marcadores pronósticos y la eficacia y seguridad de los regímenes terapéuticos, de rehabilitación y prevención. La evidencia clínica externa invalida los test diagnósticos y los tratamientos previamente aceptados y los reemplaza por otros nuevos, más potentes, más exactos, más eficaces y más seguros.

La toma de decisiones por parte de los médicos resulta difícil porque nos enfrentamos al dilema de basarla en informaciones imperfectas o inadecuadas. A esto se le llama "el principio de incertidumbre en la toma de decisiones clínicas". Según dicho principio no es posible tener una información que sea al mismo tiempo verídica y completa (Bernstein 2004). Por lo tanto nos quedan dos opciones poco atractivas: la imperfección por un lado y la inadecuación por el otro. No podemos encontrar para cada caso la información perfecta. Para la elección con sabiduría y no dejarla a la suerte, Sackett y Haynes (Sackett 1999) aconsejan los cinco pasos fundamentales de la MBE: 1) formulación de una pregunta clínica clara y precisa a partir de un problema clínico dado, es decir, formular preguntas que puedan responderse definiendo al paciente, los posibles tratamientos existentes y los mejores resultados, 2) búsqueda de la literatura de artículos originales relevantes y apropiados para el problema. La recopilación de un gran número de artículos sobre una determinada materia es algo relativamente fácil. Sin embargo, reunir las mejores evidencias disponibles de una investigación sistemática puede resultar más difícil, debido al gran número de citas irrelevantes, 3) evaluación crítica de la validez y utilidad de los artículos encontrados (Nivel de Evidencia), 4) aplicación de los resultados a la práctica clínica, es decir, la puesta en práctica de las evidencias y 5) valoración del proceso: la valoración del rendimiento es esencial para la práctica de la MBE, aunque puede tener sus problemas. La mayor dificultad reside en que la valoración se ha basado en observaciones

personales, y por tanto dicha evidencia está sujeta a sesgo. A causa de las limitaciones de la observación personal, es mejor realizar la valoración como parte de un esfuerzo conjunto.

La correcta realización de la pregunta clínica es importante para valorar las opciones terapéuticas de un determinado paciente. Una adecuada consulta y valoración de las evidencias nos ayudará a conocer los posibles resultados de dichas opciones. Para tomar una decisión, el médico simplemente debe elegir la mejor opción de una lista. A veces el proceso es muy directo; otras puede resultar más incierto. Las causas de dicha incertidumbre son básicamente dos: la primera se refiere a las llamadas zonas de vacío en las evidencias, por lo que el médico se ve obligado a interpolar. La segunda se debe a que el valor asignado a algunos de los resultados puede ser variable. La forma lógica de elegir la mejor de ellas es utilizar un árbol de decisión. En dicho árbol se incluyen todas las opciones. Las ramas que salen de ellas conducen a unos nódulos terminales que representan los posibles resultados de una determinada opción. A cada rama se le asigna un peso específico proporcional a la probabilidad de que se llegue a un determinado resultado, y a los nódulos terminales se les asignan valores que corresponden con los de cada resultado, por tanto, elegir la mejor opción terapéutica significa elegir la opción con un mayor valor esperado.

### **1.1.1 Niveles de Evidencia**

El nivel o grado de evidencia clínica es un sistema jerarquizado basado en cinco niveles.

La principal clasificación de la evidencia científica se puede obtener en su versión original a través del Oxford Centre for Evidence Based Medicine

(<http://www.eboncall.org/content/levels.html>). (Anexo A). En la revista de cirugía de la mano, The Journal of Hand Surgery, se facilita una clasificación simplificada en la sección de "normas para los autores" de su web:

(<http://journals.elsevierth.com/periodiclas/yjhsu>). De este modo, según el nivel de evidencia científica, los diseños clínicos se pueden clasificar en: Estudios Descriptivos (serie de casos, estudios de corte transversos); Estudios Observacionales (casos control, estudios de cohortes); Estudios Experimentales o Ensayos Clínicos Controlados y Aleatorizados (ECA), y finalmente, el nivel máximo de evidencia científica lo representan el Meta-Análisis (MA) y las Revisiones Sistemáticas (RS).(Rosales 2008)

### **1.1.1.1 Estudios Descriptivos**

Constituyen el nivel más bajo de evidencia científica. En este nivel se incluyen la mayoría de trabajos que se publican hoy en día o se presentan a congresos de sociedades científicas. Todos tienen un denominador común: suelen ser retrospectivos, y sólo se incluye una medición en el tiempo. Se pueden clasificar en estudios que recogen información individual (informe de un caso, serie de casos, vigilancia epidemiológica y estudios transversales), que son los más extendidos en nuestra especialidad; o de grupo (estudios ecológicos) de muy poco uso en cirugía de la mano. La opinión del experto y la observación personal tienen el nivel de evidencia más bajo.(Rosales 2008)

**1.1.1.1.1 La Opinión del experto:** la opinión del experto es la base en la que se apoya el conocimiento. Los buenos estudiantes preguntan a sus profesores y consultan los libros de texto (no los artículos de las revistas) cuando quieren comenzar a

estudiar una determinada área de conocimiento. Teniendo en cuenta que solemos aceptar las enseñanzas de los expertos en cuestiones básicas, puede resultarnos difícil rechazar sus razonamientos cuando especulen con ellos. Sin embargo, debemos aceptarlos. Una vez más, hay que recordar la jerarquía de las evidencias: la opinión del experto es mejor que la que no lo es, pero peor que la derivada de una buena investigación clínica.

**1.1.1.1.2 La observación personal:** la observación personal es la peor forma de evidencia. Podemos engañarnos con nuestras propias observaciones, incluso aunque seamos conscientes de ello e intentemos compensarlo. Algunos de los déficits de las observaciones es que por ejemplo se basan en muestras pequeñas, que son estudios sin control y que están sometidas a sesgos de recuerdo. Como señalan Redelmeier y Tversky, la gente quiere que el mundo tenga sentido e intenta dar explicaciones para lo que quiere ver, independientemente de que dichas explicaciones sean ciertas. Si un paciente es tratado y queda bien es humano que crea que se curó gracias al tratamiento, simplemente porque la curación siguió al tratamiento. Se confunden las asociaciones en el tiempo, con las relaciones causa-efecto. (Bernstein 2004)

**1.1.1.1.3 Series de casos:** la importancia de las series de casos es que pueden ser el detonante para otros estudios de mayor evidencia.

**1.1.1.1.4 Estudios de corte o transversales:** los estudios de corte o transversales son un procedimiento de observación sin seguimiento clínico

(transversal), en el que una muestra representativa de la comunidad se estudia en un momento dado. También es posible que los participantes se estudien en diferentes momentos, pero cada uno una única vez. El uso de los estudios transversales se centra fundamentalmente en el diagnóstico de salud de una comunidad; en la valoración de pruebas diagnósticas; en el análisis de validez de instrumentos de medición de resultados; y, con ciertas limitaciones, en la investigación causal sobre todo en el análisis de la ocurrencia de las enfermedades. Las ventajas de este diseño clínico son: fáciles de realizar, baratos y generalizables al basarse en la población general. El ejemplo más ilustrativo es el análisis de la prevalencia de una enfermedad o condición patológica. Como ejemplo de un estudio transversal es la valoración de la calidad de vida y discapacidad de Miembros Superiores en músicos comparada con la población general (Rosales 2005)

### **1.1.1.2 Estudios Observacionales**

Constituyen un nivel superior de evidencia científica. Se distinguen los estudios de casos y controles y los estudios de cohortes.

**1.1.1.2.1 Estudios de casos y controles:** los estudios de casos y controles normalmente son diseños clínicos de carácter retrospectivo y se dedican sobre todo al análisis causal, y a la relación de factores de riesgo con el desarrollo de una enfermedad. Los pacientes entran en el estudio en base a la variable resultado o dependiente (tener o no tener la enfermedad o condición patológica) y retrospectivamente se analiza el factor bajo estudio o variable independiente, por ejemplo, el trastorno del sueño como causa de síndrome del túnel carpiano (STC). Se

comparan de forma retrospectiva los pacientes con STC (casos) y pacientes sin STC (controles), analizando el porcentaje que existe entre los pacientes con STC y trastorno del sueño (variable independiente) y los pacientes sin STC y trastorno del sueño. Una mayor frecuencia de la variable o factor bajo estudio en el grupo de casos con respecto al control implicaría asociación causal, calculándose el "odds ratio (OR)" o razón de riesgo de exposición, un parámetro con el mismo significado que el riesgo relativo (RR), que no se puede calcular en este tipo de diseños, pero al que se aproxima cuando la enfermedad es poco frecuente. (Rosales 2008)

**1.1.1.2.2 Estudios de cohortes o longitudinales:** los estudios de cohortes o longitudinales se clasifican en tres tipos: cohortes clásico o prospectivo; cohortes histórico o retrospectivo; y cohortes bi-direccionales. Los estudios cohortes se basan en la existencia de más de una medición en el tiempo, normalmente una medición antes y otra después. Su uso se destina a la valoración de la asociación causal de un factor de riesgo con una enfermedad o bien para el análisis de resultados de un tratamiento (outcomes), o también para analizar la validez de construcción de un instrumento de medición.

Aquellos estudios se han denominado "estudios antes y después del tratamiento", con al menos 2 medidas en las variables dos veces en el tiempo. El efecto del tratamiento es analizado mediante el tamaño del efecto (ES) y la media estandarizada de la respuesta (SRM). Los resultados se pueden medir prospectivamente (estudio de cohortes clásico) o retrospectivamente (estudio de cohorte histórico). Un diseño clínico de cohorte puede también ser usado en estudios pronósticos. Siempre los pacientes entran en el diseño clínico en base al factor bajo estudio o variable independiente, expuesto o no expuesto, a diferencia de los estudios de casos-controles.

Por ejemplo, una muestra de pacientes con trastornos del sueño (expuesta) y otra muestra sin trastornos del sueño (no expuesta), en un estudio de cohorte longitudinal, se analiza la incidencia de STC en pacientes expuestos dividida entre la incidencia de STC de los no expuestos dando igual al resultado del riesgo relativo. Por último, en los cohortes bidireccionales, el diseño comienza como los cohortes históricos. El investigador viaja hacia atrás de forma retrospectiva, recolecta las medidas preoperatorias, pero a diferencia de los cohortes históricos, cita a los pacientes en su unidad y recoge de forma prospectiva las mediciones post-operatorias. (Rosales 2008)

### **1.1.1.3 Ensayos Clínicos Controlados Aleatorizados**

Denominados también estudios experimentales. Los ensayos clínicos aleatorizados (ECA) son todos estudios de cohortes prospectivos en los que se comparan dos tratamientos y se quiere saber cuál es el mejor; o bien se quiere analizar la eficacia de un tratamiento, y para ello se compara con un control o placebo. Siempre es deseable tener en los ECA un grupo experimental (recibe el tratamiento bajo estudio) y un grupo control (recibe un placebo o tratamiento sin eficacia terapéutica; o bien, el tratamiento clásico empleado anteriormente). Todos los pacientes incluidos en el ECA tienen la misma probabilidad "p" de recibir el tratamiento y pertenecer al grupo experimental, que la probabilidad "1-p" de pertenecer al grupo control, es decir, cualquiera de los pacientes que entra en un ECA tiene la misma probabilidad de recibir un tratamiento que el otro.

Las características principales de los ECA son: 1: asignación al azar de los pacientes a uno u otro grupo de tratamiento, lo cual reduce los sesgos de selección y los de confusión. El sesgo de selección se reduce al no permitir que el investigador sea quien asigne la intervención, y el de confusión se reduce por la tendencia que tiene el

azar a repartir por igual en los grupos que se comparan los distintos factores que pueden influir en el resultado final. 2: son controlados, es decir, existen al menos dos grupos de tratamiento, con el fin de reducir la influencia de ciertos factores que puedan dificultar la valoración de la eficacia del tratamiento experimental y 3: deben estar enmascarados, y a ser posible a doble ciego, es decir, mientras el estudio no finaliza, ni el médico ni el paciente conocen la asignación de tratamientos con el fin de evitar sesgos y minimizar el efecto "Hawthorne" (influencia positiva en los resultados debido al conocimiento previo de estar incluido en un estudio). El enmascaramiento puede ser: a) a simple ciego, son aquellos en los que los sujetos que entran en el ECA no saben si están recibiendo el tratamiento en estudio o el placebo; b) a doble ciego, cuando ni los pacientes ni los médicos que tratan y siguen los pacientes conocen la asignación de los tratamientos; c) a triple ciego cuando existe un analista independiente que tampoco conoce la asignación de pacientes.

La importancia de los ECA es que constituyen el diseño clínico de mayor evidencia científica que se realiza en la investigación clínica habitual, pues por encima de estos estudios sólo existe el meta-análisis y la revisión sistemática.

#### **1.1.1.4 El Meta-Análisis y La Revisión Sistemática**

El Meta-Análisis (MA) y la Revisión Sistemática (RS) constituyen los estudios de mayor evidencia científica en investigación clínica.

El Meta-Análisis (Huque 1998), es un análisis estadístico que combina o integra los resultados de varios ECA considerados por el analista como combinables. El término Revisión Sistemática es más apropiado para definir cualquier revisión de un conjunto de datos que usa claramente un método y criterios definidos. Así, una RS

puede incluir un MA; apreciaciones o revisiones de un solo ECA, y otros recursos de evidencia.

Los puntos esenciales del MA son: a) permite una evaluación más objetiva de la evidencia, lo cual puede conducir a la resolución de incertidumbres y discrepancias; b) puede disminuir la probabilidad de resultados falsos negativos y, por lo tanto, prevenir demoras innecesarias en la introducción de tratamientos eficaces en la práctica clínica; c) permite evaluar la hipótesis formulada a priori acerca de los efectos del tratamiento en subgrupos de pacientes; d) permite explorar y a veces explicar la heterogeneidad entre los resultados de los estudios; e) se pueden generar preguntas de investigación prometedoras que se evaluarán en estudios futuros, y f) se puede calcular el tamaño muestral necesario en estudios de forma precisa.

Existen muy pocos Meta-Análisis que hayan sido publicados en revistas científicas de cirugía de la mano, y algunos de ellos no son realmente Meta-Análisis.

El Meta-Análisis es probablemente el diseño clínico de mayor evidencia científica donde el investigador combina los resultados de estudios previos publicados, preferiblemente ECA o bien Estudios Observacionales, combinables entre sí, con el fin de calcular una estimación global del efecto del tratamiento.

## **1.1.2 Finalidad de la Medicina Basada en la Evidencia**

En esencia, la Medicina Basada en la Evidencia aporta un marco conceptual nuevo para la resolución de problemas clínicos concretos mediante un proceso de búsqueda sistemática, evaluación crítica y aplicación de los hallazgos de la investigación a la toma de decisiones clínicas; es un medio para afrontar la existencia de una enorme información científica y en evolución continua, la exigencia de ofrecer la

máxima calidad asistencial y la limitación de recursos destinados a la atención sanitaria. Es un método de trabajo que se puede considerar de excelencia, al combinar en uno, el concepto de alta calidad documental, la unificación de las fuentes de información y el acceso a éstas mediante artículos a texto completo. El mejor ejemplo de esta metodología es la Biblioteca Cochrane plus (<http://www.update-software.com/clibplus.asp>) a la que podemos acceder de forma gratuita por la página del Ministerio de Sanidad y Consumo.

### **1.1.2.1 La Colaboración Cochrane**

La Colaboración Cochrane es una organización sin ánimo de lucro que reúne a un grupo de investigadores de Ciencias de la Salud de más de 11500 voluntarios en más de 90 países. Su nombre proviene de Archie Cochrane, un epidemiólogo británico, que creía que las intervenciones en salud que han demostrado su efectividad deben ser gratuitas. La Colaboración Cochrane fue fundada en 1993 por Iain Chalmers en el Reino Unido para facilitar las revisiones sistemáticas de ensayos controlados aleatorios en todos los ámbitos de la atención de la salud. Chalmers argumentaba que los escasos recursos en el cuidado de la salud podían ser eficientemente manejados sólo si nosotros conocíamos qué tratamientos eran efectivos y cuáles no. Denunció también la falta de accesibilidad a informaciones de este tipo y destacó la necesidad de llevar a cabo una revisión periódicamente actualizada. En respuesta a este reto nació en Oxford, en 1992, el Cochrane Center, y en octubre de 1993 un total de setenta y siete personas de once países cofundaron la Colaboración Cochrane. Esta organización defiende que las evidencias científicas sobre los efectos de un determinado tratamiento son solamente una parte de los elementos a considerar en una toma de decisión óptima. Se deben

atender las particularidades de cada caso concreto, como son las necesidades del paciente, la disponibilidad de recursos, la pericia del clínico, las prioridades de cada situación, e incluso el conocimiento del propio paciente sobre su estado. Recientemente, se han revisado todos los Meta-Análisis y Revisiones Sistemáticas incluidos en la base de datos Cochrane relacionados con la cirugía de la mano, concluyendo que de 470 Revisiones Sistemáticas, sólo existían 14 revisiones relacionadas con la cirugía de la mano (5 sobre fractura distal de radio, 4 sobre STC, 1 sobre antibióticos, 1 sobre rehabilitación de tendones flexores, 1 sobre fracturas del cuello del 5º metacarpiano, 1 sobre dedo en martillo y 1 sobre rizartrrosis). Todos estaban comprometidos a nivel de metodología y heterogeneidad clínica de los estudios incluidos (Schädel-Höpner 2008)

La Medicina Basada en la Evidencia no se restringe a los ensayos clínicos aleatorizados y meta-análisis (Sackett 1996). Implica perseguir la mejor evidencia externa con la que responder a nuestras preguntas clínicas. Así, para conocer acerca de la exactitud de una prueba diagnóstica, necesitamos encontrar estudios transversales apropiados de pacientes con sospecha clínica de presentar la alteración en cuestión, no un ECA y para una cuestión sobre pronóstico, necesitamos adecuados estudios de seguimiento de pacientes agrupados en un punto precoz y uniforme del curso clínico de su enfermedad. Y a veces, la evidencia que requerimos vendrá de las ciencias básicas como la genética y la inmunología. Cuando se hacen preguntas sobre tratamientos debemos tratar de evitar los enfoques no experimentales, puesto que éstos llevan habitualmente a conclusiones sobre eficacia falsamente positivas. Una de las fuentes más importantes de "pseudoinformación pseudocientífica" procede de las compañías farmacéuticas. El hecho de que en nuestro medio sean las principales fuentes de soporte financiero de la investigación y la formación de los médicos, deriva en una información sesgada, dirigida a intereses comerciales, y en la que lo "científico" y lo "promocional"

es a menudo indistinguible. Por lo tanto, debemos buscar Ensayos Clínicos Aleatorizados y sobretodo Revisiones Sistemáticas de varios ECA. De este modo conseguiremos que nos informen sobre el tratamiento más apropiado y será menos probable que nos confundan, es por lo que ambos se han convertido en el "gold standar" para juzgar si un tratamiento proporciona más beneficio que daño. Sin embargo, algunas cuestiones sobre tratamientos no requieren ensayos aleatorizados (por ejemplo intervenciones con éxito para problemas de otro modo fatal) o no se puede esperar a que se realicen esos ensayos. Si no se ha llevado a cabo ningún ECA para el problema de nuestro paciente, seguimos el camino hasta la siguiente mejor evidencia externa y trabajamos a partir de ella.

A pesar de sus orígenes antiguos, la Medicina Basada en la Evidencia continúa siendo una disciplina relativamente joven cuyo impacto positivo sólo está empezando a ser validado y continuará evolucionando. Esta evolución será reforzada conforme numerosos programas de pregrado, postgrado y de educación médica continuada, adopten y se adapten a las necesidades de quienes las aprendan.

### **1.1.3 ¿Cuál es la calidad en las publicaciones de los Ensayos Clínicos Aleatorizados en la Extremidad Superior?**

Es más que reconocido en la Medicina Basada en la Evidencia, que los ECA son estudios importantes que pueden llegar a determinar la efectividad de diferentes opciones de tratamiento. El resultado de un ECA tiene generalmente un gran impacto, ya que al compararlo con otro diseño de ECA, se disminuye el riesgo de sesgo. Sin

embargo, el tipo de diseño clínico no asegura que el estudio sea de alta calidad. Para los lectores de un artículo clínico, la calidad del estudio es importante, ya que de ella depende si el resultado final justifica que el lector lo aplique en su práctica clínica. Es indicador de la validez interna de un ECA el control de sesgos: la selección del paciente, el análisis de datos, el manejo de abandonos, cómo se han valorado los resultados. Se entiende por sesgo el error sistemático en un estudio epidemiológico que produce una estimación incorrecta de asociación entre la exposición y la enfermedad. Los principales sesgos a los que nos enfrentamos son el sesgo en la selección, y el sesgo en la información u observación. El sesgo de selección hace referencia a cualquier error que deriva del proceso de identificación de la población a estudiar. Se puede cometer, al seleccionar el grupo control, al seleccionar el espacio muestral donde se realizará el estudio, por pérdidas en el seguimiento o por la presencia de una supervivencia selectiva. El sesgo de información u observación incluye cualquier error sistemático en la medida de la información sobre la exposición a estudiar o los resultados. Las fuentes más frecuentes de este sesgo son: instrumento de medida no adecuado, criterios diagnósticos incorrectos, omisiones, imprecisiones en la información, errores en la clasificación y errores introducidos por los cuestionarios o las encuestadoras. La principal razón para la randomización o aleatorización en los ensayos clínicos comparando 2 intervenciones, es prevenir el sesgo de selección que puede resultar en diferencias sistemáticas entre los pacientes sometidos al tratamiento. Esta es la base para validar el resultado de cualquier diferencia observada en el tratamiento que se evalúa. La asignación aleatoria de los pacientes incluidos en un ensayo clínico según los criterios de inclusión y exclusión, se entiende como la probabilidad de dar a cada paciente la misma oportunidad de recibir un tratamiento y de asegurarnos de no saber cuál ni a quién ha sido. La importancia de publicar correctamente los ECA está siendo

reconocida mundialmente. Una publicación inadecuada disminuye la certeza de las conclusiones y la fuerza del mensaje que se quiere transmitir con el estudio. (Christina Gummesson 2004)

Para mejorar la calidad de los ECA se publicó en 1996 la guía CONSORT (The Consolidated Standards for Reporting Trials) así como diferentes escalas para evaluar la calidad de los ECA, siendo la escala de Jadad la más usada. Existen dos revisiones de la guía CONSORT, una en el 2001 y la última en el 2010. La declaración CONSORT 2010 consta de una lista de comprobación de 25 ítems que se consideran críticos para determinar la calidad de un ECA. Por tanto, su comprobación deberá incluirse en todo manuscrito de un ensayo clínico, que se envíe para su publicación junto con un diagrama que ilustre el flujo de individuos participantes a lo largo del ensayo. Aunque estas herramientas pueden orientar para valorar cualquier ensayo clínico, se centran en el diseño más común: el de dos grupos paralelos con aleatorización de individuos. Otros diseños, como los ensayos con aleatorización de conglomerados (en los que se aleatorizan grupos o "conglomerados" de individuos), pueden requerir información adicional y existen extensiones de CONSORT específicas para ellos. La traducción al castellano de la declaración CONSORT 2010 se realizó mediante un proceso de traducción y retrotraducción. (Cobos A.2011). La lista de los 25 ítems está disponible en la web. (<http://www.consort-statement.org/Downloads/checklist.doc>).

La escala de Jadad o sistema de puntuación de calidad de Oxford, es un procedimiento para evaluar de manera independiente la calidad metodológica de un ensayo clínico. Es la evaluación más ampliamente utilizada en el mundo. (Jadad AR. 1996). La escala de Jadad considera sólo aquellos aspectos relacionados con los sesgos referidos a la aleatorización y el enmascaramiento (conocido como doble ciego), que impiden que los pacientes y el propio investigador tengan acceso al objeto del

tratamiento y a la descripción de las pérdidas de seguimiento. Es un cuestionario validado, sencillo y rápido de aplicar que consta de 7 preguntas. Mediante éstas preguntas puede realizarse una valoración de la calidad de un ensayo clínico: 1) ¿El estudio se describe como aleatorizado (o randomizado)? Sí: 1 punto, No: 0 puntos. 2) ¿Se describe el método utilizado para generar la secuencia de aleatorización y este método es adecuado? Sí: 1 punto, No: 0 puntos. 3) ¿Es adecuado el método utilizado para generar la secuencia de aleatorización? Sí: 0 puntos. No: -1 punto. 4) ¿El estudio se describe como doble ciego? Sí: 1 punto No: 0 puntos. 5) ¿Se describe el método de enmascaramiento (o cegamiento)? Sí: 1 punto No: 0 puntos. 6) ¿Es adecuado el método de enmascaramiento o cegamiento)? Sí: 0 puntos. No: -1 punto. 7) ¿Hay una descripción de las pérdidas de seguimiento y los abandonos? Sí: 1 punto No: 0 puntos. Este cuestionario da una escala que va de 0 a 5 puntos, de manera que a mayor puntuación mejor calidad metodológica tiene el ensayo clínico evaluado. Se considera un ECA riguroso con 5 puntos, y un ECA de pobre calidad si su puntuación es inferior a 3 puntos.

Otro aspecto importante para evaluar la eficacia de una intervención es cómo se ha realizado la valoración de resultados y si cubre los aspectos del efecto del tratamiento esperado. El estado funcional está relacionado con la condición de salud pero la condición de salud no permite predecir el estado funcional. La organización mundial de la salud define salud, no como ausencia de enfermedad, sino como un estado completo de bienestar físico, mental y de salud.

En este sentido en el año 2001 se aprobó la International Classification of Functioning, Disability and Health, conocida como ICF, como marco para proporcionar un medio de clasificar la función y detrimento asociado a la condición de salud. Se entiende el concepto de discapacidad, no como un resultado final, sino como

el nivel de discapacidad que presenta un individuo y que podemos medir. Es por tanto una clasificación de la salud y de las condiciones relacionadas con la misma, y proporciona una manera unificada de clasificar los componentes de la salud, al entender ésta desde un punto de vista multidimensional. Estas condiciones de la salud están relacionadas con la estructura y función corporal, con la actividad y sus limitaciones, con la participación de la persona en áreas de la vida y con factores medioambientales que afectan a sus experiencias. La ICF no mide la discapacidad al ser un modelo descriptivo, pero proporciona un lenguaje común de los conceptos de medida de salud y discapacidad.

Siguiendo estos principios, Christina Gummesson e Isam Atroshi en el año 2003, realizaron un estudio acerca de cuál era la calidad de los ECA de estudios sobre patologías del miembro superior publicados durante un periodo de 10 años (desde 1992 a 2002) en 4 revistas: *Journal of Hand Surgery (American volume)*, *Journal of Hand Surgery (British and European volume)*, *Journal of Bone and Joint Surgery (American volume)*, and *Journal of Bone and Joint Surgery (British volume)*. Un total de 92 artículos fueron incluidos, siendo las patologías más frecuentemente estudiadas, la fractura distal de radio y el síndrome del túnel carpiano. Como resultado de esta valoración sistemática, se obtuvo que sólo el 43% de los estudios publicados como ECA relacionados con patologías del miembro superior verdaderamente eran ECA. 23% de las publicaciones eran definidas como ECA pero el método de randomización descrito era inapropiado, y en el 34% restante, el método de randomización no fue descrito, lo cual hacía imposible saber si eran verdaderos estudios randomizados. En comparación con otro estudio similar, que valoró la calidad de los ECA entre los años 1992-1996, se ha visto un aumento, estadísticamente significativo, en la mejoría de la

calidad de los ECA en relación a la aplicación de los criterios de la Guía CONSORT publicados en 1996. (Gummesson 2004)

Pero la publicación de estos datos de baja calidad en revistas de nuestra especialidad, no eran una sorpresa. Muchos han criticado la falta de ensayos clínicos en el campo de la cirugía, y cuando éstos se llevaban a cabo, su pobre rigor científico y falta de metodología originaban muchas críticas y eran ridiculizadas (Horton 1996). Como lo describió Horton, editor de la revista médica británica *The Lancet*, en su comentario "Surgical Research or Comic Opera: Question, but few Answers", criticando la falta de rigor científico en las publicaciones de cirugía. Pero, ¿son estas críticas justificadas?. Son justificadas cuando el investigador realiza un estudio sin metodología alguna, perdiendo el tiempo tanto él como el de los pacientes y, lo peor, haciendo conclusiones sobre los efectos de un tratamiento que conducirán a errores. Pero por otro lado, la crítica puede estar no justificada. Los ensayos clínicos quirúrgicos son diferentes de los ensayos que comparan una medicación con otra. Los procedimientos quirúrgicos son invasivos y es difícil randomizar a los pacientes. Por ejemplo, un ECA para probar la efectividad de la liberación del túnel carpiano puede requerir que a un grupo se le haga una falsa cirugía en la cual sólo se realice una incisión cutánea. Estos ensayos, aunque atractivos, pueden no ser aceptables ni para el paciente ni para el cirujano. Otra consideración importante es el coste y la duración de estos estudios. En muchas sociedades quirúrgicas estos ensayos están prohibidos debido al aumento de coste por aumento en el personal para su realización. La mayoría de los ECA quirúrgicos requieren un largo periodo de seguimiento ya que sabemos que un resultado a los 6 meses puede no ser igual que a los 3 años, lo que a su vez debido a la rápida evolución en los avances técnicos y cambios en las técnicas quirúrgicas, lo que es un buen resultado puede haber quedado obsoleto a los 5-10 años de un ensayo

clínico.(Chung 2004). La principal razón para la randomización en los ECA es la prevención de sesgos de selección que pueden dar como resultado diferencias sistemáticas entre los grupos de estudio. La asignación del paciente a cada grupo debe ser totalmente al azar. Es inapropiado el uso de datos como la fecha de nacimiento, el día que fue a la consulta, etc. Como en cirugía el procedimiento a doble ciego no se lleva a cabo en la mayoría de estudios, se modificó la escala de Jadad introduciendo el concepto de simple ciego, en el que sólo el investigador o el paciente, pero no ambos, es consciente de la naturaleza de la intervención. Por ejemplo, se consideró adecuado que el paciente intervenido de síndrome del túnel carpiano llevara un guante para cubrir el área quirúrgica durante la valoración o se hicieran idénticas incisiones en todos los pacientes aunque las intervenciones fueran diferentes y sobre todo lo que mejoraba la calidad del ensayo era que la persona encargada de la valoración de resultados no supiera nada sobre el grupo en estudio.(Gummesson 2004)

### **1.1.4 Y ante estos datos, ¿qué podemos hacer para mejorar nuestros trabajos de investigación clínica en cirugía de la mano?**

Es importante mantener presente dos preguntas en toda investigación clínica: la importancia de los antecedentes en la generación de la pregunta de investigación u objetivos ("el por qué?"), y el método ("¿cómo?"). La descripción de los resultados ("lo descubierto") y la discusión de los mismos ("la interpretación de los resultados" ) son menos importantes. (Rosales 2015)

En la introducción (background) donde nos preguntamos "el por qué", se debe exponer cuál es el conocimiento científico actual sobre el tema objetivo del estudio (qué se conoce), cuál es el vacío en el conocimiento (qué no se conoce), generar la pregunta de investigación e hipótesis del trabajo (sintetiza la pregunta científica a la que se pretende dar respuesta con el trabajo de investigación), para finalizar con la parte fundamental de este apartado que es el objetivo o propósito del estudio de investigación (qué se quiere conocer). La pregunta de investigación representa la incertidumbre del investigador sobre algún fenómeno que ocurre en la población y debe ser: importante (por la gravedad y frecuencia), interesante (a la luz de los conocimientos actuales), que tenga respuesta (con los medios disponibles) y realizable (en un tiempo en concreto).

Los estudios de investigación, en función de la cuestión fundamental que el investigador quiere contestar con su trabajo, se pueden clasificar en estudios sobre pruebas diagnósticas o un test, estudios sobre la etiología o causalidad de una enfermedad, estudios sobre el pronóstico de una enfermedad, y estudios sobre la eficacia-efectividad de una terapia o tratamiento (outcomes). (Haynes 1983)

Los estudios sobre pruebas diagnósticas, son aquellos en los que el investigador se plantea el uso o aplicación de una nueva prueba diagnóstica en sus pacientes. El proceso de generalización de la pregunta de investigación comienza con el conocimiento, a partir de la bibliografía, sobre si existe evidencia sobre la comparación a ciegas e independiente con otro test o prueba diagnóstica denominada "estándar de oro". Los estudios sobre la causalidad, son los que intentan determinar la etiología o factores causales de una enfermedad. En este tipo de estudios se pueden incluir los trabajos destinados a conocer la ocurrencia de una enfermedad o patología en términos de incidencia o prevalencia, pues constituyen el primer peldaño para sospechar y generar preguntas de investigación sobre factores causales que determinen la

distribución y frecuencia de una patología o enfermedad. Los estudios sobre la eficacia y efectividad de un tratamiento medicamentoso o quirúrgico se conocen a nivel internacional como "outcomes" o de valoración de resultados, y el propósito de los mismos va encaminado a distinguir entre un tratamiento útil de aquel inútil o incluso perjudicial.

A menudo se confunden los términos de eficacia, efectividad, y eficiencia. Estaremos hablando de eficacia cuando se mide el resultado o el efecto de decidir en sanidad en el marco experimental de un ECA. Hablaremos de efectividad cuando se haya medido el resultado, el efecto de esa decisión, mediante estudios observacionales. Y entenderemos como eficiencia la medida del coste de la decisión en sanidad. Lo más eficiente es lo de mejor calidad al menor coste. Los ensayos o los análisis de coste-efectividad serían el método para determinar la eficiencia de una determinada decisión. Normalmente, los estudios de coste-efectividad se establecen para distinguir, para una patología determinada, de entre aquellos tratamientos considerados buenos (de probada eficacia) cuál es el mejor (más coste - efectivo), conociéndose el gasto por incremento de "QALY" (unidad de calidad de vida ajustada a la expectativa de edad en años) (Quality-Adjusted Life-Years).

La sección de Material y Método ("¿cómo?") constituye la parte más importante de cualquier trabajo de investigación clínica. De ella dependerá si los resultados obtenidos pueden aceptarse como evidencia científica. Debe definir la población estudio, el diseño clínico y los instrumentos y medidas empleados. En la población estudio, hay que especificar claramente cuándo un paciente es "caso" mediante unos criterios clínicos y test complementarios. De este modo, nos aseguramos de que cada caso es diagnosticado de la misma forma. Sólo la revisión bibliográfica previa, en relación a la pregunta de investigación que queremos contestar, nos dará aquellos

criterios de inclusión más recomendables para definir nuestra población estudio. La definición de un paciente como "caso" y su inclusión en la población estudio dependerá de los criterios diagnósticos que empleemos. Así llegamos a la conclusión de que el diagnóstico en Medicina, bajo la visión científica, se convierte en un cálculo de probabilidades, y ello dependerá del grado de certeza de los criterios clínicos y pruebas complementarias que usemos para establecer dicho diagnóstico. El grado de certeza de cualquier diagnóstico médico va a depender de la sensibilidad (capacidad de una prueba diagnóstica en detectar los verdaderos positivos (VP) entre todos los que verdaderamente están afectados de la enfermedad:  $VP/VP+FN$  (FN: Falsos Negativos), y de la especificidad (capacidad de una prueba diagnóstica en detectar los verdaderos negativos (VN) de entre todos los pacientes que no están afectados de la enfermedad:  $VN/VN+FP$  (FP: Falsos Positivos). Cuanto más sensible sea una prueba más probable es que detecte a las personas o sujetos con la enfermedad. Es decir, un resultado negativo casi excluiría la posibilidad de que el paciente tenga la enfermedad. Esto deriva en el hecho de que una prueba muy sensible provoca pocos FN. Además si la prueba es muy específica indica mayor probabilidad de que las personas sin la enfermedad sean así consideradas por la prueba diagnóstica. Las pruebas muy específicas son a menudo usadas para confirmar la presencia de la enfermedad. Para establecer a un paciente como "caso" necesitamos saber qué nivel de sensibilidad y especificidad deberíamos usar en una prueba diagnóstica o criterio. Para ello se utilizan las "Curvas ROC" (ROC: Received Operating Characteristic). Las curvas de ROC son gráficos en los que se enfrentan en el eje de ordenadas la Sensibilidad (S), y en el eje de abscisas el complemento de la Especificidad (1-E), o lo que es lo mismo los FP. En cirugía de la mano, las curvas ROC han sido utilizadas para determinar la sensibilidad y especificidad de diferentes test clínicos y neurofisiológicos en el diagnóstico del

síndrome del túnel carpiano, para detectar en la población general, aquellos síndromes del túnel carpiano graves que potencialmente necesitan descompresión quirúrgica, para calcular la diferencia mínima de importancia clínica (MCID) que se puede medir a nivel de la gravedad de síntomas en el STC etc.. De la misma manera, los criterios de exclusión nos van a definir quién está fuera de nuestra población estudio. Los criterios de exclusión intentan disminuir los factores que pueden influir en los resultados de la investigación, y que dependen de la población diana. El riesgo potencial de categorizar casos es que ocurran sesgos de información. Dos aspectos importantes de los sesgos son la confiabilidad (consistencia interna y reproductibilidad) y la validez, parámetros que se desarrollan en el apartado de Instrumentos de Medición.

Una vez hemos establecido nuestra pregunta de investigación y definido la población estudio, se ha de definir el diseño clínico, piedra fundamental del material y método. El nivel de evidencia de nuestro estudio de investigación dependerá del diseño clínico (Anexo A).

Por último, los Instrumentos y Medidas empleados nos permiten evaluar el efecto de una intervención o tratamiento. Tradicionalmente las medidas que suelen usarse para valorar resultados en cirugía de la mano se basan fundamentalmente en el estado funcional físico y estructural, como en radiografías, medida de la fuerza de agarre y el rango de movilidad (Gummenson 2004). Recientemente, disponemos de medidas estandarizadas en base a la opinión del paciente, denominadas instrumentos PRO (Patient Reported Outcomes), como por ejemplo el Disabilities of the Arm, Shoulder and Hand (DASH), Michigan Hand Questionnaire (MHQ), Patient Evaluation Measure (PEM) y el Patient Rated Wrist Evaluation (PRWE).

La información obtenida del análisis de los datos clínicos, junto a la valoración de resultados sobre la base de la opinión del paciente, puede proporcionar un espectro

multidimensional del resultado de tratamiento denominado en la literatura anglosajona como "Outcome of Care", mediante el cual puede ser evaluada la calidad del servicio sanitario y el valor de dichos servicios para el individuo y la sociedad.

## 1.2 Movimiento Outcomes

Arnold Relman, editor jefe del *The New England Journal of Medicine*, en 1990, definió el "Outcomes Movement" como "La tercera revolución en medicina". Una nueva era sin precedentes, de crecimiento dirigido a la valoración de resultados, al análisis de la eficacia y a la seguridad en la calidad asistencial. Los defensores del nuevo movimiento de medir resultados en la práctica médica, pronosticaban grandes beneficios sociales, incluyendo la mejor información tanto a médicos como pacientes, mejoría de las guías clínicas y más sabias decisiones para los "compradores de la salud". Pero estos esfuerzos, ¿merecían la pena?. Como Elwood dijo se necesita "un sistema nervioso central que nos ayude a hacer frente a las complejidades de la medicina moderna ". (Epstein, 1990)

Por lo menos tres factores importantes promulgaron el desarrollo del Movimiento Outcomes. El primero la necesidad de contención de costes; permitiría la valoración de la efectividad de diferentes intervenciones eliminando los gastos innecesarios y sería una parte vital del sistema de sanidad, dirigido no sólo a mejorar la calidad de la asistencia sino a detectar su deterioro. La realización de "Guías" a tal fin, dado su coste tanto económico como de tiempo, serían sólo justificables para procedimientos comunes o para aquellos que tienen un importante impacto en costes o resultados. El segundo factor es un renovado sentido de la competencia y el tercero y quizás el más importante, el estudio en la valoración de resultados de diferentes procedimientos médicos cuando existen enormes diferencias en su uso en distintas áreas geográficas, lanzando preguntas como si en áreas donde hay un mayor uso de un procedimiento, esto refleja un aumento del coste de manera innecesaria, o si por el contrario es sub-óptimo en áreas donde no se realiza el mismo procedimiento. (Epstein,

1990). Wennerg en 1987 comparó los días de hospitalización de 2 hospitales situados demográficamente en áreas similares y encontró diferencias significativas en el uso de ciertos procedimientos y en cambio la tasa de mortalidad era la misma. Este estudio demostraba que el incremento del uso en los servicios médicos no mejoraba los resultados. (Wennerg 1987)

En respuesta a estas preocupaciones se aprobó la Agencia para la Investigación y Calidad de los Servicios de Salud "The Agency for Healthcare Research and Quality, (AHRQ)". La misión de la AHRQ era identificar estudios de investigación que mejoraran directamente el resultado de los pacientes, y que produjeran cambios en nuestra práctica clínica creando guías para el uso apropiado de los servicios médicos. La AHRQ desarrolló un modelo para mostrar el impacto de los estudios de investigación basado en 4 niveles. El nivel 1 de impacto, describe un estudio que presenta una nueva información clínica de un tratamiento ya existente o una nueva técnica o un nuevo instrumento, que tiene el potencial de cambiar la decisión clínica actual. Por ejemplo, una nueva placa distal de radio presentada en un estudio prospectivo multicéntrico. No se compara con un grupo control, pero demuestra que la placa es segura y efectiva. El nivel 2 de impacto, describe el uso de datos basados en la evidencia de órganos de gobierno, para justificar la creación de una nueva política o programa. El nivel 3 de impacto, describe el cambio en una decisión o acto del médico y/o paciente a consecuencia de datos específicos de investigación, guías clínicas o normas establecidas. El nivel 4 de impacto, describe un estudio o tendencia histórica que muestra una mejoría de los resultados (clínicos, económicos, de calidad de vida y satisfacción) como resultado directo de una intervención específica que ha sido comparada con un grupo control.

El movimiento outcomes empezó con la esperanza de reducir costes eliminando procedimientos innecesarios, mejorando la difusión de la información a pacientes y médicos, mejorando la salud y función de los pacientes y desarrollando guías para la práctica clínica. Una revisión sistemática llevada a cabo por la AHRQ en el año 2000, sin embargo, encontró pocos cambios en la práctica clínica y política: muchos de los objetivos del movimiento outcomes no se habían cumplido. En un estudio similar (K.Chung 2006), se valoró el nivel de impacto de los artículos científicos publicados en el Journal of Hand Surgery desde el año 1988 (comienzo del movimiento outcomes) al año 2004. El 92% de los artículos revisados tenían un nivel de impacto 1, es decir, no contribuyen directamente a cambiar la decisión en la práctica clínica o política ni en el resultado final de salud del paciente. Por ejemplo, en un estudio retrospectivo de 2 grupos de pacientes que fueron intervenidos de síndrome del túnel carpiano de manera abierta o por endoscopia. En la valoración de resultados no se encontraron diferencias respecto al alivio de síntomas, tiempo en reincorporarse al trabajo o complicaciones. Pocos estudios, sólo un 9% se clasificaron como nivel 4. Un ejemplo lo tenemos en el uso de ultrasonidos para el diagnóstico de lesiones sólidas o quísticas en la mano y muñeca respecto a la impresión clínica. El ultrasonido fue muy preciso (83%) en diagnosticar lesiones en mano y muñeca en comparación con la primera impresión clínica (54%), esta clara ventaja hace que se cambie la práctica clínica al respecto. La ausencia de estudios con nivel 2 y nivel 3, indica que la valoración de resultados en cirugía de la mano no ha progresado lo suficiente como para cambiar la decisión y la práctica tanto del médico como del paciente o de crear guías clínicas que tengan implicaciones en la política sanitaria. Es de señalar que el número de artículos con nivel 4 aumentó desde 1988 (ninguno) al 2004 (el 11% de los estudios eran nivel 4), por lo tanto estamos en la buena dirección de mejorar. Sin embargo, mucho trabajo queda

por hacer respecto a los estudios con nivel 2 y 3. Sería por ejemplo de gran potencial para cambiar el modo de actuar del médico, si existiera una guía para tratar el dolor crónico de muñeca.

Existe un interés creciente en nuestra especialidad por el empleo de instrumentos de valoración del estado de salud para la evaluación de resultados. Durante muchos años de investigación se han valorado los resultados en base a datos puramente objetivos de la exploración y el examen clínico realizados por el médico o cirujano, pero éstos no aportaban información alguna sobre cómo un problema localizado en la mano, impacta en otras dimensiones de la salud, como por ejemplo, la funcionalidad social, la vitalidad o incluso a nivel emocional o en la salud mental. Estas dimensiones de la salud relacionadas con la calidad de vida son los parámetros que preocupan al paciente, los que necesitamos mejorar con nuestros tratamientos, y los que debemos utilizar a la hora de valorar la eficacia de dichos tratamientos. La práctica médica se debe considerar un éxito sólo si contribuye a mejorar el bienestar del paciente. Este principio ha constituido una de las más importantes fuerzas externas que han impulsado la entrada del "Outcomes Movement" (Epstein 1990) ( Resultado final o efecto) para designar el impacto que un determinado proceso o tratamiento tiene sobre el bienestar del paciente, cómo modificar el estado de salud o la calidad de vida. A diferencia de los parámetros clínicos, la determinación del resultado final es un proceso centrado en la valoración que el propio paciente hace de su salud. Este movimiento se basa en la superioridad del conocimiento científico sobre cualquier otro conocimiento médico. Así, se ha establecido un verdadero enfrentamiento entre la medicina-ciencia basada en la evidencia científica, y la medicina-arte, basada en la experiencia.

La idea de resultado final no es nueva. Fue introducida por el cirujano ortopédico Codman (1934), con el propósito de mejorar la eficiencia del cuidado de los

pacientes. Esta idea quedó relegada durante gran parte del siglo XX por varios motivos. Por una parte, la disponibilidad de enormes recursos económicos y el entusiasmo depositado en supuestos avances médicos restaron interés al estudio del coste y a la eficacia real de estos últimos. Por otra parte, la aparición de nuevos métodos de diagnóstico y tratamiento dificultó la comparación de los resultados obtenidos, en un mismo proceso, con diferentes sistemas de tratamiento y evaluación. (Keller RB 1996). Codman estableció el concepto de "Outcome" que radicaba en comprender qué resultados producían los tratamientos médicos y quirúrgicos en la recuperación y desarrollo de las actividades diarias de los pacientes. Sin embargo, desde hace tan sólo 15 años la comunidad ortopédica, especialmente la anglosajona, viene participando de forma activa en este llamado movimiento-valoración de resultados. Gartland, una de las figuras principales de dicho movimiento, puso de manifiesto las deficiencias evidentes en la bibliografía médica. Durante los últimos 10 años el Outcomes Committee of the American Academy of Orthopaedic Surgeons (Comité de Resultados de la Academia Americana de Cirujanos Ortopédicos) viene ofreciendo cursos cada año sobre las técnicas empleadas en la evaluación de resultados.

La mejora de los sistemas de información, la voluntad política y social de contener el gasto sanitario y una cierta preocupación por la metodología y la demostración de evidencias científicamente constatadas, han reavivado durante las dos últimas décadas el interés de la comunidad médica en general y la ortopédica en particular por el estudio de resultados finales.

Actualmente se continúa ofertando un tratamiento médico-quirúrgico en base a la experiencia, pero la experiencia aislada no implica buena praxis médica; sólo la evidencia científica implica buena práctica médica. Este es el objetivo último del movimiento "outcomes", conocer si realmente lo que hacemos en nuestra práctica

médica diaria es realmente eficaz basándonos en la evidencia científica. Por tanto, de un sistema clásico, paternalista y bidimensional de la sanidad, constituido por los profesionales sanitarios y el pagador de la sanidad (compañía privada, sistema público o sistema de "tercer sector") se ha evolucionado -dentro del movimiento outcomes- a un sistema competitivo tridimensional donde se ha incluido la medición científica de la opinión del paciente a la hora de valorar la eficacia de cualquier intervención sanitaria. Se ha demostrado que estas medidas científicas de la opinión del paciente son más sensibles a los cambios clínicos tras el tratamiento, que los datos de la exploración física realizados por el propio médico, a la hora de medir resultados. De esta manera, se ha introducido el uso de Instrumentos de Medición de la Calidad de Vida, basados en la opinión del paciente para valoración de resultados en medicina

## 1.3 Instrumentos de Salud (Instrumentos PRO)

En los últimos 30 años, se ha consolidado el área de investigación en los resultados percibidos por los pacientes, especialmente en el ámbito de la medida de la Calidad de Vida Relacionada con la Salud (CVRS). Gran parte de los esfuerzos se han consagrado a la construcción de instrumentos de medida con propósitos diversos. Los más de 1270 instrumentos de resultados percibidos por los pacientes identificados en el ámbito internacional y los más de 350 concebidos para su uso en la población española varían notablemente en cuanto a su contenido, los métodos seguidos para su obtención, sus propiedades métricas, las necesidades para su aplicación y su uso efectivo. Se han realizado numerosas propuestas para sistematizar la evaluación de estos instrumentos, destacando en el ámbito español, la Red de Investigación en Resultados y Servicios Sanitarios (Red IRYSS), una de cuyas líneas de investigación incluye la evaluación sistemática de los instrumentos de medida de CVRS disponibles para su uso en España, que se concreta en el proyecto ePRO (Evaluación de Resultados Percibidos por los Pacientes). El proyecto ePRO ha consistido en el desarrollo y aplicación de un procedimiento estandarizado de evaluación de instrumentos de medida. (J. M. Valderas 2005)

Los instrumentos de valoración del estado de salud representan un intento de determinar con mayor precisión la percepción que cada persona tiene de su bienestar o malestar independientemente de los parámetros clínicos que el médico determine.

Medir la satisfacción de un paciente ante un resultado es una tarea muy complicada. Los pacientes pueden variar en su nivel de satisfacción teniendo los mismos síntomas y estado funcional, incluso el propio paciente puede cambiar su percepción a lo largo del tiempo. Hay variedad también a la hora de asumir un

determinado problema, por miedo a lo desconocido o por no someterse a tratamientos invasivos, de aquí la importancia, de valorar el estado funcional y emocional de manera individualizada.

Los instrumentos que usamos para medir resultados son multidimensionales. Hay un componente anatómico y un componente funcional. El anatómico es el que estamos acostumbrados a usar, son las medidas que obtenemos con el uso del goniómetro, monofilamentos de Semmes-Weinstein y aparatos para medir fuerza de pinza y puño. Sin embargo, es importante entender que la funcionalidad del paciente va paralela a su anatomía. Por ejemplo, un músico tendrá mayor discapacidad que otra persona ante la misma patología en la mano. (Amadio 2001), (Rosales 2005)

Debemos comprender claramente los diferentes temas que intervienen en la evaluación de los resultados clínicos y funcionales. Los resultados clínicos corresponden a lo que tradicionalmente medimos, es decir, los denominados parámetros clínicos: tasas de infección, amplitud de movimiento, fuerza, etc. Frecuentemente estos parámetros se combinan para establecer escalas de valoración. Los parámetros clínicos son útiles pero no siempre presentan una correlación perfecta con el grado de bienestar del paciente. Generalmente los pacientes que presentan menor rango de movilidad articular se encuentran peor que los que tienen rango de movilidad completo, pero no siempre es así. Más aún, aunque nuestro tratamiento consiga restablecer un rango de movilidad completo, eso no asegura el restablecimiento del bienestar del paciente. (Donabedian 1989)

Brenner et al. realizaron una distinción práctica entre medición de resultados a nivel proximal y a nivel distal. Con resultados proximales se refiere a aquellos que son más cercanos, caracterizados por la expresión de la enfermedad en sí, como el dolor y parestesias nocturnas del síndrome del túnel carpiano. Resultados menos proximales se

refiere, y a veces olvidados, a funciones físicas como por ejemplo la imposibilidad de tocar el piano durante varias horas porque aparecen síntomas del túnel carpiano. Los resultados distales son aquellos que están más alejados de la enfermedad, como la satisfacción o el bienestar. Brenner señaló la importancia de establecer esta continuidad en la valoración de resultados, demostrando que cuando las medidas proximales eran más severas, en las medidas distales se producía una mejoría mayor como resultado final tras tratamiento, en este caso del síndrome del túnel carpiano.(Brenner 1995)

El primer requisito para valorar el impacto de un determinado procedimiento sobre la calidad de vida es utilizar instrumentos validados para la población a estudiar, teniendo en cuenta variaciones culturales y lingüísticas. Estos datos se obtienen del paciente mediante el uso de cuestionarios validados y se centran en el desarrollo de las funciones y actividades de cada paciente como un todo, partiendo desde el punto de vista de éste.

### **1.3.1 Tipos de Instrumentos de Salud**

Existen dos grandes grupos de instrumentos de valoración del estado de salud: genéricos y específicos. Lo ideal es emplear simultáneamente un cuestionario genérico (que mida la salud global) y uno específico (específicamente designado para un tipo de paciente o patología concretos).

#### **1.3.1.1 Instrumentos Genéricos**

Los instrumentos genéricos, como el conocido cuestionario de salud SF-36, son cuestionarios que tratan de medir todas las dimensiones importantes de la salud en

relación a la calidad de vida, y pueden ser usados en cualquier tipo de afección o patología independientemente de la condición subyacente. Suelen tener menor respuesta o sensibilidad para detectar cambios de importancia clínica tras tratamiento, pero permiten la comparación entre diferentes enfermedades, patologías o condiciones. Por ejemplo, cuando el SF-36 es utilizado en una población específica, como en artrodesis segmentarias del carpo, puede que no detecte una importante mejoría tras tratamiento comparado con un instrumento específico de un área o territorio funcional, como el cuestionario DASH (Disability of Arm, Shoulder and Hand) ; sin embargo, nos permite conocer el incremento general en la calidad de vida comparado con otros tipos de tratamiento como el conservador, o incluso con los tratamientos de otras patologías distintas de la mano. Ello es sumamente importante para los organizadores de la sanidad a la hora de hacer un mejor uso de los recursos sanitarios de una determinada comunidad, región o país. Así, en un estudio (Rosales 2002) se objetivó que en Canarias el detrimento de la calidad de vida a nivel de la dimensión de la salud de funcionalidad física en una población afecta de STC era casi similar a la de una población de enfermos coronarios estables. Y, hoy en día se destina más dinero de fondos públicos en campañas de prevención de la enfermedad coronaria comparado a lo que se destina en una afección tan localizada en la mano como el STC.

### **1.3.1.2 Instrumentos Específicos**

Los instrumentos específicos se centran en problemas asociados a una condición o patología específica, grupos o poblaciones de pacientes, o áreas de función. Presentan una mayor respuesta o sensibilidad para detectar cambios de importancia clínica tras tratamiento; pero no permiten comparación de resultados a través de diferentes

patologías o condiciones. Dentro de este tipo de cuestionarios son de destacar por ejemplo: el cuestionario específico para el síndrome del túnel carpiano (Instrumento CTS) (Instrumento de Levine) (Brigham and Women's Hospital, Boston Carpal Tunnel Instrument) y el DASH (Disability of Arm, Shoulder and Hand), (Institute for Work & Health, AAOS, ASSH 1996) que es un instrumento específico de un área de función, y nos sirve para medir la gravedad de los síntomas y el estado funcional del miembro superior.

### **1.3.2 ¿Cómo se desarrolla un Instrumento?**

El procedimiento de construcción de un instrumento de medida de resultados percibidos por el paciente es complejo y de naturaleza interactiva. Ello hace que normalmente los atributos de un instrumento se evalúen en distintos artículos científicos. Los criterios usados en el proyecto ePRO se basan en la propuesta del comité científico del Medical Outcomes Trust, de este modo los instrumentos PRO se desarrollan de acuerdo a unos principios psicométricos establecidos. Los pasos que se siguen son: el desarrollo de las cuestiones que se van a preguntar, la selección del formato, la simplificación de las preguntas, la confiabilidad o reliability, la validación (validity) y el análisis de la respuesta (responsiveness) (Marc F.Swiotkowski, 1999), la interpretabilidad de las puntuaciones, la carga para el entrevistador y para el entrevistado y la disponibilidad, características de formatos alternativos al estándar y el proceso de adaptación transcultural. (J.M.Valderas 2005)

1. Las cuestiones o preguntas se obtienen a través de expertos en el tratamiento del proceso que se evalúa, mediante otros cuestionarios o, como sería lo ideal, a través de los pacientes que sufren el problema o problemas que se están

estudiando. El cuestionario puede tener un formato de respuesta SI/NO, un formato de escala o un formato de escala analógico visual donde el paciente hace una marca en un segmento lineal establecido. Hay diferentes ventajas y desventajas para cada uno de estos formatos.

2. La validez (validity) de un instrumento de medición se define como la capacidad que presenta un instrumento para medir el concepto que propone medir. De una manera algo menos genérica se podría entender como el grado en que las puntuaciones obtenidas con el instrumento reflejan adecuadamente la calidad de vida relacionada con la salud del individuo o de las poblaciones a las que se aplica. Existen tres maneras de acumular evidencia sobre la validez de un instrumento; a) Validez de contenido: evidencia de que el contenido dominante o dominio del instrumento de medición es apropiado y relativo al uso que se le pretende dar. Suele evaluarse mediante la participación de paneles de expertos y de pacientes. b) Validez de criterio: implica la recogida simultánea de datos junto con el uso de un modelo establecido para determinar cómo se relacionan los datos recogidos a través del nuevo cuestionario con los datos estándar. La falta de medidas de referencia ha hecho que, la evaluación de este tipo de validez se haya aplicado casi exclusivamente al estudio de versiones reducidas de instrumentos, empleando el instrumento original como estándar. c) Validez de construcción. Es una de las más utilizadas en nuestro grupo de trabajo. En ella, el investigador intenta establecer una relación lógica entre las medidas obtenidas con un instrumento con otras medidas de probada validez. (Atroshi 1998). Tradicionalmente se ha diferenciado entre validez de constructo convergente o estudio de correlación con medidas similares, y validez de

constructo divergente, orientado a la identificación de falta de asociación con medidas dispares o de diferencias en las puntuaciones entre grupos diferenciados o extremos según criterios clínicos o epidemiológicos. Para su correcta aplicación resulta imprescindible la formulación de hipótesis a priori sobre los patrones de relación esperados. Se ha propuesto que valores de correlación mayores o iguales de 0.60 se corresponderían con una asociación de elevada intensidad y que los menores de 0.30 indicarían falta de asociación. (JM Valderas 2005). Así los autores establecieron por primera vez el uso del análisis de la respuesta para valorar la validez de construcción de la versión española del DASH. El equipo investigador estableció la siguiente hipótesis: el instrumento DASH aplicado a una población específica de STC presentará una respuesta mayor a la que presentaría un instrumento genérico como el SF-36; y por otro lado, esta respuesta del DASH sería menor a la que pudiera presentar un instrumento específico de la enfermedad como el cuestionario de Levine (Rosales 2001). Cuando los investigadores concluyen que hay una verdadera diferencia entre los dos grupos de tratamiento (rechazando la hipótesis nula) o no hay diferencia entre los grupos (aceptando la hipótesis nula) hay que estar seguros de tener una alta probabilidad de estar en lo cierto. Hay que minimizar el riesgo de cometer un error tipo I o  $\alpha$  (rechazar la hipótesis nula cuando esta es correcta, es decir no existe diferencia entre los grupos) y un error tipo II o  $\beta$  (se acepta que no hay diferencia cuando sí existe diferencia entre los grupos). Un estudio es susceptible de generar un error tipo II cuando el tamaño de la muestra es pequeño o la varianza del estudio es muy grande. Para concluir que no hay diferencia entre los grupos de estudio el investigador debe de establecer a priori tanto el error  $\beta$  o el poder  $(1-\beta)$  como la probabilidad de que una verdadera

diferencia sea correcta cuando ésta es presente. Cuando un poder es  $>80\%$ , indica que hay menos del 20% de perder una diferencia real. Estudios con un poder de  $< 80\%$  tiene mayor probabilidad de cometer un error  $\beta$  y concluir erróneamente que no hay diferencia entre grupos. El fallo de no tener en cuenta el estadístico  $\beta$  puede llevar a dos indeseables consecuencias: una de tipo ético, por la pérdida de tiempo y esfuerzo al paciente, y segundo de tipo académico al invalidar un ensayo clínico y no permitir que otros investigadores usen las alternativas de tratamiento para otros estudios. Por lo tanto cuando un estudio concluye que no hay diferencia entre los dos grupos de tratamiento ("estudios negativos"), es esencial determinar si el estudio tenía suficiente poder para encontrar una diferencia clínicamente significativa. Insuficiente poder excluye un adecuado resultado de un tratamiento eficaz y resulta en un error tipo II, una conclusión errónea de aceptar la hipótesis nula. En su estudio, Kevin Chung, revisó 39 "estudios negativos" de la revista *The Journal of Hand Surgery* y se comprobó que 32 (82%) de los artículos tenían un poder de menos del 80% . (KChung 1998)

3. La confiabilidad (reliability) de un Instrumento de medición se ha definido tradicionalmente como el grado en que un instrumento está libre de error aleatorio o ausencia de error sistemático. Las fuentes de error aleatorio se han buscado en las respuestas proporcionadas a los diferentes ítems que componen una misma escala en una misma administración del instrumento (consistencia interna), entre las distintas administraciones de un mismo instrumento a la misma población (fiabilidad test-retest) o entre distintos observadores o administradores (fiabilidad interobservador). La confiabilidad abarca por lo

tanto dos conceptos muy importantes: la consistencia interna y la reproductibilidad. Se entiende por consistencia interna la habilidad que tienen los ítems de una escala de un instrumento para medir un concepto común a todos ellos. Para su evaluación se utiliza el coeficiente alfa de Cronbach. Se han propuesto como estándares mínimos apropiados valores de alfa de Cronbach de 0.70 para realizar comparaciones entre grupos y de 0.90-0.95 para realizar comparaciones individuales, porque de otra manera el intervalo de confianza de las puntuaciones individuales resulta tan amplio que el uso del instrumento resultaría ineficaz para una buena consistencia interna (Advisor Scientific Committee of Medical Outcome Trust 1998). Se define como reproductibilidad a la estabilidad de un instrumento a través del tiempo. Se analiza mediante el test-retest, mediante dos mediciones separadas en un tiempo en la que se espera que la enfermedad no haya cambiado. Con frecuencia, sólo es posible asegurar esta estabilidad entre administraciones inmediatas muy repetidas. Cuando se utilizan medidas de la exploración clínica (ej. fuerza, rango de movilidad..), el mismo principio debe ser aplicado en cuanto a la fiabilidad, y esta viene determinada por analizar los errores de observación e inter-observador, que se calculan mediante el estadístico de concordancia o "Kappa" o el coeficiente de correlación intraclase, por tratarse de medidas continuas. Los estándares mínimos recomendados son los mismos que para la valoración de la consistencia interna.

4. Respuesta de un instrumento de medición, o sensibilidad al cambio (responsiveness): se conoce como respuesta a la sensibilidad de una medida o

instrumento de medición para detectar cambios de importancia clínica en el tiempo. El conocimiento de cuáles son las medidas más sensibles para valorar resultados en una población estudio específica es un paso fundamental en el campo de investigación clínica. El análisis de la respuesta o sensibilidad para detectar cambios va a depender de muchos factores: así por ejemplo, los tamaños poblacionales de un estudio sobre eficacia de un tratamiento o técnica quirúrgica en el campo de la cirugía de la mano van a depender de la sensibilidad o respuesta de los instrumentos o medidas empleados. Cuanto más sensibles o mayor respuesta tengan las mediciones que vamos a usar en nuestro estudio, menor tamaño poblacional necesitaremos para encontrar diferencias significativas. Por otro lado, el análisis de la respuesta nos dará la información sobre cuáles son las medidas de elección a emplear para esa población específica y para el tiempo estimado de seguimiento o valoración. En caso contrario, podríamos llegar a conclusiones erróneas en nuestro estudio. Podemos observar malos resultados tras tratamiento debido a que a lo mejor estamos usando medidas que no son sensibles para detectar cambios de importancia clínica, cuando en realidad el tratamiento ha sido eficaz. La respuesta se mide en base al "Tamaño del Efecto" ( $ES = \text{effect size}$ ) que se calcula dividiendo la media del cambio (de antes del tratamiento a después) por la DS (Desviación Estándar) pre-tratamiento; y la "Media Estandarizada de la Respuesta" ( $SRM = \text{Standardized Response Mean}$ ) que se mide en base a la media del cambio dividida por la DS del cambio. Estos estadísticos constituyen la unidad de medida de la mejoría clínica tras tratamiento y su significado está en que nos dicen qué diferencias estadísticas significativas tienen, o no, importancia clínica. Así, en nuestra línea de investigación, se demostró que las medidas del DASH, y

de otros instrumentos de salud, presentan una mayor capacidad para detectar los cambios tras tratamiento (mejor respuesta), comparándolos con las medidas de la exploración clínica, que veníamos usando durante estos últimos 50 años en cirugía de la mano, a la hora de valorar los resultados del tratamiento quirúrgico del STC.(Rosales 2009). La idea de seleccionar o desarrollar instrumentos sensibles es entendida en conjunto con la diferencia mínima clínicamente importante (MCID), que ha sido bien establecida en algunos de los instrumentos que se usan en la extremidad superior, identificando el cambio más pequeño que debe ser considerado clínicamente significativo en el resultado de un instrumento. La sensibilidad al cambio también está vinculada al efecto techo-suelo. Estas características se refieren a la capacidad de un instrumento de identificar cambios en los extremos de las escalas del instrumento. Por ejemplo, si incrementos en la mejoría existen por encima de lo que la escala puede medir, el instrumento tiene un efecto techo debido a que estos cambios no son identificados. De manera inversa, si la escala tiene un efecto suelo, puede no ser capaz de medir valores por debajo de cierto nivel. (M.Bdalamente 2013)

5. Interpretabilidad: es el grado en que es posible asignar un significado comprensible a las puntuaciones de un instrumento.
  
6. Carga para el entrevistado y entrevistador: se define como el tiempo, esfuerzo, necesidades y exigencias derivadas de la administración de un instrumento. Se mide en relación con consideraciones de índole operativa, como el tiempo que

se tarda en administrar el cuestionario, requisitos mínimos para una administración efectiva, capacidad de lectura y comprensión necesaria por parte del entrevistado.

7. Formatos de administración: las formas básicas de administración de los instrumentos incluyen la autoadministración (supervisada o no), la administración mediante personal entrenado (telefónica o en persona), y la administración a personas próximas distintas del paciente (padres, cuidadores etc).
  
8. La comparabilidad transcultural: la adaptación cultural y lingüística es el octavo criterio propuesto por el comité científico del Medical Outcomes Trust. Este criterio es de suma importancia ya que la mayoría de los instrumentos de medida de que se dispone son adaptaciones de cuestionarios originalmente desarrollados en otros países. Con el fin de garantizar la aplicabilidad de estos instrumentos a otras poblaciones en general, y a la española, en particular, es preciso alcanzar la equivalencia conceptual y lingüística con el original. Para el proceso de adaptación transcultural se considera recomendable realizar al menos 2 traducciones independientes del instrumento del idioma original al de adaptación y al menos una traducción inversa (retrotraducción) de la anterior de nuevo al idioma original. Es muy conveniente la participación en los estudios de adaptación de población bilingüe en ambos idiomas y con un cierto grado de conocimiento de ambas culturas.

### **1.3.3 Instrumentos de Salud Genéricos en la Valoración de Resultados en el Miembro Superior**

Los especialistas en patología de la mano y miembro superior somos afortunados ya que disponemos de instrumentos que valoran tanto el estado funcional como físico. La Sociedad Americana de Terapeutas de la Mano ha publicado gran diversidad de recomendaciones tales como la evaluación del edema, la cicatriz, la circulación, fuerza de puño, fuerza muscular, rango de movilidad, sensibilidad, destreza, dolor y capacidad funcional.

Como ocurre en las demás áreas de nuestra especialidad, la adecuada valoración de los resultados obtenidos con diferentes tratamientos en la patología del miembro superior requiere la utilización de varias herramientas. La exploración física y las escalas de puntuación clásicas proporcionan una información extremadamente valiosa, sin embargo, la valoración de resultados no se considera completa si no se valoran otros aspectos, como la intensidad del dolor, la satisfacción y la calidad de vida percibida por el paciente.

La mayor parte de la patología del miembro superior se manifiesta en forma de dolor. Por ello, una parte fundamental de la valoración de resultados la constituye la valoración del dolor residual y satisfacción global que el paciente pueda presentar finalizado un determinado tratamiento o intervención, utilizando por ejemplo escalas analógico visuales (EVA).

La exploración física continúa siendo una herramienta indispensable para la valoración de los resultados. Ciertos parámetros, como la movilidad articular en diferentes planos permiten una monitorización del cambio a lo largo del tiempo.

Los cuestionarios de calidad de vida percibida por el paciente se han convertido en una herramienta esencial en la valoración de resultados. Como dijimos previamente, estos cuestionarios se pueden clasificar en dos grandes grupos en base al tipo de información recolectada: genéricos y específicos. Lo ideal al realizar un estudio es combinar ambos.

En general, los instrumentos genéricos de salud juegan un papel menor como valoración de resultados en la extremidad superior. Su papel más importante es el de valorar ideas, de hecho, como instrumento primario es poco utilizado, sin embargo, junto a un instrumento específico, proporciona la mejor valoración de resultados en la calidad de vida de un paciente que se puede realizar.

De los diferentes cuestionarios genéricos existentes, probablemente los de mayor utilidad en el campo de la extremidad superior, como en muchos otros, son el SF-36, y el EuroQol-5D (EQ-5D) de más reciente aparición. Nos centraremos en estos dos Instrumentos de Salud Genéricos, ya que son los que hemos empleado en nuestras hipótesis de trabajo.

### **1.3.3.1 Cuestionario SF-36**

El cuestionario de salud SF-36 (Thirty - Six Items Short Form Health Survey) fue desarrollado a principio de los noventa, en Estados Unidos, para su uso en el Estudio de los Resultados Médicos (Medical Outcomes Study, MOS). Es una escala genérica que proporciona un perfil del estado de salud y es aplicable tanto a los pacientes como a la población general. Ha resultado útil para evaluar la calidad de vida relacionada con la salud (CVRS) en la población general y en subgrupos específicos, comparar la carga de muy diversas enfermedades, detectar los beneficios en la salud

producidos por un amplio rango de tratamientos diferentes y valorar el estado de salud de pacientes individuales. Sus buenas propiedades psicométricas, que han sido evaluadas en más de 400 artículos, y la multitud de estudios ya realizados, que permiten la comparación de resultados, lo convierten en uno de los instrumentos con mayor potencial en el campo de la CVRS. (Alonso 1995).

El Cuestionario de Salud SF-36 está compuesto por 36 preguntas (ítems) que valoran los estados tanto positivos como negativos de la salud. Se desarrolló a partir de una extensa batería de cuestionarios empleados en el MOS, que incluían 40 conceptos relacionados con la salud. Para crear el cuestionario, se seleccionó el mínimo número de conceptos necesarios para mantener la validez y las características operativas del test inicial. El cuestionario final incluye 8 escalas, que representan los conceptos de salud empleados con más frecuencia en los principales cuestionarios de salud, así como los aspectos más relacionados con la enfermedad y el tratamiento. Los 36 ítems del instrumento cubren las siguientes escalas: función física, rol físico, dolor corporal, salud general, vitalidad, función social, rol emocional y salud mental. Adicionalmente, el SF-36 incluye un ítem de transición que pregunta sobre el cambio en el estado de salud general respecto al año anterior. Este ítem no se utiliza para el cálculo de ninguna de las escalas pero proporciona información útil sobre el cambio percibido en el estado de salud durante el año previo a la administración del SF-36.

Hay 2 versiones del cuestionario en cuanto al periodo recordatorio: la "estándar" (4 semanas) y la "aguda" (1 semana).

El cuestionario está dirigido a personas mayores de 14 años de edad y preferentemente debe ser autoadministrado, aunque también es aceptable la administración mediante entrevista personal y telefónica. La consistencia interna no

presentó diferencias entre los cuestionarios autoadministrados y los aplicados mediante entrevista. (Ware 1993)

La traducción al español del cuestionario siguió un protocolo común en los países participantes en el proyecto internacional de adaptación del cuestionario original, el Internacional Quality of Life Assessment (IQOLA) (Alonso 1998). Los investigadores de proyecto IQOLA se esforzaron por intentar mantener la comparabilidad de las diferentes versiones al adaptar culturalmente los ítems. Por ejemplo, la actividad de jugar al golf de la versión original se tradujo por ir en bicicleta en Italia y Holanda, caminar por el bosque o por jardines en Suiza, y caminar durante más de 1 h en España.

Las escalas del SF-36 están ordenadas de forma que a mayor puntuación mejor es el estado de salud. Puntuación 100 sería el mejor estado de salud y puntuación 0 sería el peor estado de salud. En caso de que falte información, si se han contestado al menos el 50% de los ítems de una escala, los autores recomiendan sustituir cualquier ítem ausente por el promedio de los ítems completados de ésta. En caso contrario (más del 50% de ítems no contestados), la puntuación de dicha escala no se debería calcular.

Aunque el cuestionario SF-36 puede ser cumplimentado en un periodo que oscila entre 5 y 10 minutos, en determinados contextos puede representar demasiado tiempo, por ello se desarrollaron el SF-12 y SF-8, que han demostrado ser una alternativa útil al SF-36 cuando se pretende medir la salud física y mental en general y el tamaño de la muestra es elevado (500 individuos o más). Para estudios más pequeños o cuando se pretende estudiar una o más escalas del SF-36 por separado, es preferible utilizar el SF-36. (Gandek 1998)

Es necesario señalar que el SF-36 no incluye algunos conceptos de salud importantes, como los trastornos del sueño, la función cognitiva, la función familiar o la

función sexual. Ello es especialmente relevante cuando se diseñan estudios en que los pacientes sufren enfermedades o efectos secundarios de tratamientos que pueden afectar a alguna de estas escalas o conceptos de salud.

En la siguiente tabla se describen las diferentes dimensiones, el número de ítems y el significado de puntuaciones altas y bajas ( Ware 1993).

Tabla 2: Contenido de las escalas del SF-36

Dimensión	Nº de ítems	"Peor" puntuación (0)	"Mejor" puntuación (100)
Función física	10	Muy limitado para llevar a cabo todas las actividades físicas, incluido bañarse o ducharse, debido a la salud	Lleva a cabo todo tipo de actividades físicas incluidas las más vigorosas sin ninguna limitación debido a la salud.
Rol Físico	4	Problemas con el trabajo u otras actividades diarias debido a la salud física	Ningún problema con el trabajo u otras actividades diarias debido a la salud física
Dolor corporal	2	Dolor muy intenso y extremadamente limitante	Ningún dolor ni limitaciones debidas a él
Salud General	5	Evalúa como mala la propia salud y cree posible que empeore	Evalúa la propia salud como excelente
Vitalidad	4	Se siente cansado y exhausto todo el tiempo	Se siente muy dinámico y lleno de energía todo el tiempo
Función social	2	Interferencia extrema y muy frecuente con las actividades sociales normales, debido a problemas físicos o emocionales	Lleva a cabo actividades sociales normales sin ninguna interferencia debido a problemas físicos o emocionales
Rol emocional	3	Problemas con el trabajo y otras actividades diarias debido a problemas emocionales	Ningún problema con el trabajo y otras actividades diarias debido a problemas emocionales
Salud Mental	5	Sentimiento de angustia y depresión durante todo el tiempo	Sentimiento de felicidad, tranquilidad y calma durante todo el tiempo
Ítem de transición de salud	1	Cree que su salud es mucho peor ahora que hace un año	Cree que su salud general es mucho mejor ahora que hace un año

Hay diferentes procesos-enfermedades que afectan a la mano y extremidad superior que tienen un gran impacto y por tanto afectan al bienestar y salud general del paciente, como por ejemplo la mano reumática, lesiones del plexo braquial, mutilaciones severas o manos traumáticas. La función emocional puede verse afectada

en estos casos y hay muy pocos instrumentos con subescalas que midan este aspecto. (Badalamente 2013)

La versión española del SF-36 es uno de los instrumentos genéricos más utilizados en el territorio nacional, tanto en estudios descriptivos que miden el impacto sobre la CVRS en distintas poblaciones de pacientes como para la evaluación de intervenciones terapéuticas. Además se ha utilizado como referencia en la validación de nuevos instrumentos de medición. (<http://www.sf-36.org>). (Anexo E)

### **1.3.3.2 Cuestionario Euroqol-5D**

El cuestionario EQ-5D ( Euro Qol Group, 1990), es un instrumento genérico de medición de la calidad de vida relacionada con la salud, que puede utilizarse tanto en individuos relativamente sanos (población general) como en grupos de pacientes con diferentes patologías. El propio individuo valora su estado de salud, primero en niveles de gravedad por dimensiones (sistema descriptivo), y luego en una escala visual analógica (EVA) de evaluación más general. Un tercer elemento del EQ-5D es el índice de valores sociales que se obtiene para cada estado de salud generado por el instrumento. El sistema descriptivo contiene cinco dimensiones de salud (movilidad, cuidado personal, actividades cotidianas, dolor/malestar y ansiedad/depresión) y cada una de ellas tiene tres niveles de gravedad (sin problemas, algunos problemas o problemas moderados y problemas graves). En esta parte del cuestionario el individuo debe marcar el nivel de gravedad correspondiente a su estado de salud en cada una de las dimensiones, refiriéndose al mismo día que cumplimente el cuestionario.

En cada dimensión del EQ-5D, los niveles de gravedad se codifican con 1 si la opción de respuesta es "no tengo problemas"; con un 2 si la opción de respuesta es "algunos o moderados problemas"; y con un 3 si la opción de respuesta es "muchos problemas". La combinación de los valores de todas las dimensiones genera números de 5 dígitos, habiendo 243 combinaciones-estados de salud- posibles, que pueden utilizarse como perfiles.

La segunda parte del EQ-5D es una EVA vertical de 20 centímetros, milimetrada, que va desde 0 (peor estado de salud imaginable) a 100 (mejor estado de salud imaginable). En ella, el individuo debe marcar el punto en la línea vertical que mejor refleje la valoración de su estado de salud global en el día de hoy. El uso de la EVA proporciona una puntuación complementaria al sistema descriptivo de la autoevaluación del estado de salud del individuo. El índice de valores de referencia para cada estado de salud se obtiene a partir de estudios en población general o en grupos de pacientes en los cuales se valoran varios de los estados de salud generados por el EQ-5D utilizando una técnica de valoración como el time trade-off. El índice oscila entre el valor 1 (mejor estado de salud) y el 0 (la muerte), aunque existen valores negativos, correspondientes a aquellos estados de salud que son valorados como peores que la muerte.

El EQ-5D puede ser autoadministrado, o bien realizando una entrevista telefónica.

El EQ-5D ha sido utilizado de distintas maneras en atención primaria, como por ejemplo para medir el efecto del tratamiento en pacientes con claudicación intermitente, como indicador de salud en pacientes con dolor lumbar o en pacientes con osteoartritis de rodilla, resultando en todos ser sensible a los cambios en el estado de salud debidos al tratamiento.

La ventaja más destacable del EQ-5D es que se trata de un instrumento muy corto y sencillo de rellenar. El tiempo de administración es de aproximadamente 2-3 minutos y la sencillez de su uso repercute positivamente en la cantidad y calidad de los datos recogidos, con menor número de respuestas perdidas o equivocadas. Incluso se ha comprobado que funciona relativamente bien en niños a partir de los 6 años.

Desde su adaptación en España, el EQ-5D ha sido probado en numerosos estudios que permitieron medir sus propiedades psicométricas (validez, fiabilidad y sensibilidad al cambio). (Anexo H)

### **1.3.4 Instrumentos de Salud Específicos en la Valoración de Resultados en la Extremidad Superior**

Las lesiones en la mano y extremidad superior son las más frecuentes de nuestro cuerpo. Se estima que 16 millones de lesiones en la extremidad superior ocurren cada año, lo que da a lugar a 16 millones de bajas laborales, y esto a su vez a 10 billones de dólares de coste total. En 1983, Robert A. Chase, a favor del Comité Socioeconómico de la Sociedad Americana de Cirugía de la Mano recomendó el desarrollo de un sistema universal de valoración de resultados de discapacidad en el miembro superior. No es de sorprender, por lo tanto, que la extremidad superior sea una de las áreas en las que más se ha avanzado en el uso de cuestionarios específicos para la valoración de resultados. Los instrumentos específicos se centran en problemas asociados a una condición o patología específica, grupos o poblaciones de pacientes, ó áreas de función. Presentan una mayor respuesta o sensibilidad para detectar cambios de importancia clínica tras

tratamiento; pero, no permiten la comparación de resultados a través de diferentes patologías o condiciones.

El interés creciente en la valoración de resultados por áreas específicas para cada región anatómica ha dado lugar a cuestionarios para determinados tipos de patologías como el Shoulder Pain and Disability Index (SPADI), Western Ontario Osteoarthritis of the Shoulder Index (WOOS), Rotator-Cuff Quality of Life (Rc-QOL), Oxford Shoulder Scores (OSS), el Patient-Rated Wrist-Hand Evaluation (PRWHE), y el cuestionario Brigham del síndrome del túnel carpiano (CTS).

El cuestionario más empleado para la valoración global de la extremidad superior es el cuestionario DASH (Disabilities of the Arm, Shoulder and Hand), seguido del Michigan Hand Questionnaire (MHQ) y el cuestionario CTS.

En un estudio se observó que en un periodo de doce meses en la revista Journal of Hand Surgery, los instrumentos de medida más utilizados han sido los siguientes: el DASH en 31 artículos (n=31), el MHQ en 8 artículos (n=8), el CTS en 3 artículos (n=3), el PRWHE en 2 artículos (n=2) y la escala de intolerancia al frío en 5 artículos (n=5). (Calfee 2012)

Nos centraremos en el estudio de los instrumentos específicos MHQ, DASH, QuickDASH y CTS, por su importancia en la extremidad superior.

### **1.3.4.1 Instrumento Michigan Hand Outcomes**

El "Michigan Hand Outcomes Questionnaire" es un instrumento diseñado para pacientes con diversas patologías de la mano. Después de su desarrollo preclínico se estructuró un cuestionario de 37 ítems distribuidos en seis subescalas: Funcionamiento

global de la mano, Actividades de la vida diaria, Dolor, Trabajo, Apariencia Estética de la mano y Satisfacción del paciente con la función de la mano.

El cuestionario fue sometido a un análisis de confiabilidad, consistencia interna y validez. (K.Chung 1998). La confiabilidad se demostró con el test - retest y consistencia interna, usando el coeficiente de correlación de Spearman para el test-retest con resultados desde 0.81 para la subescala estética hasta 0.97 para la subescala de actividades de la vida diaria. La consistencia interna fue alta, con valores de alfa de Cronbach, entre 0.86 para la subescala de dolor y 0.97 para la escala de actividades de la vida diaria. La correlación entre subescalas mostró evidencia de validez de constructo. Respecto a la validez discriminante de la escala estética se encontró una diferencia significativa ( $p = 0.0012$ ) entre los pacientes con síndrome de túnel del carpo y con artritis reumatoide. Se concluyó que el MHQ valora la funcionalidad de la mano, posee una alta correlación con otras escalas, es fácil de usar, bien aceptado y confiable.

El MHQ se ha aplicado en un amplio espectro de condiciones que afectan a la mano; en pacientes con reconstrucción del pulgar por transferencia libre, transferencia microvascular de múltiples dedos, intervenciones en síndrome del túnel del carpo, traumatismos en niños, reconstrucción articular en pacientes con artritis reumatoide y quemaduras de la mano. También se ha utilizado como referente para la validación de otras escalas y se ha comparado con otros instrumentos en patologías definidas.

Tiene una correlación alta con los resultados de otras escalas. Se ha encontrado que el MHQ posee validez convergente con otros instrumentos, especificidad para la mano y versatilidad al compararlo con escalas similares. La inclusión en el MHQ de una escala que mide la apariencia estética de la mano es una única e interesante componente ya que a menudo no se tiene en cuenta en pacientes con lesiones en la mano y es de suma importancia.

El MHQ ha sido traducido a múltiples idiomas y existe una versión abreviada , el Brief MHQ.

### **1.3.4.2 Instrumento DASH**

El cuestionario más empleado para la valoración global de la extremidad superior es el cuestionario DASH (Disabilities of the Arm, Shoulder and Hand), desarrollado conjuntamente por el Institute for Work and Health (IWH) y la American Academy of Orthopedic Surgeons (AAOS, 1994), para medir discapacidad y síntomas en relación con los MMSS y sus afecciones músculo-esqueléticas.

El DASH es el resultado de un esfuerzo multidisciplinar. El primer asunto por resolver era si tenía sentido que un único instrumento fuera capaz de valorar todo el miembro superior. Fueron seleccionados un conjunto de ítems de 20 cuestionarios utilizados para medir la función en la mano, muñeca, codo, hombro y antebrazo. Los ítems fueron clasificados de acuerdo con los criterios de la Organización Mundial de la Salud en grupos, según los síntomas y discapacidad. Muchas preguntas de aspecto general se convirtieron en específicas; por ejemplo, ¿tiene usted dolor? se convirtió en ¿tiene usted dolor en su mano?. Las preguntas sobre discapacidad, por ejemplo, ¿puede usted hacer un nudo? o ¿puede alcanzar algo por encima de su cabeza?, nos permiten localizar la disfunción en la mano, muñeca u hombro. Pocas cuestiones había para valorar selectivamente el codo. Finalmente, preguntas como ¿es usted capaz de seguir trabajando?, no eran específicas de una región anatómica sino que valoraban al paciente como un todo. (Amadio 2001)

El pool de cuestiones que se usaron para desarrollar el DASH fue muy extenso, llegando a recogerse hasta 800 preguntas en todo el mundo. 500 pacientes realizaron el

cuestionario en todo el mundo, todos con patologías del miembro superior, desde pacientes con artritis reumatoide, epicondilitis lateral, a un dedo en resorte. El análisis estadístico de los resultados permitió que se llegara a los 30 ítems que hoy constituye el DASH original con el módulo opcional de deporte / música.

La distinción entre los conceptos "discapacidad" y "función" es importante. La "discapacidad" es informada por el paciente y puede variar de manera importante dependiendo de lo que midamos, como la fuerza o el movimiento. Ha habido un fuerte crecimiento en la valoración de lo que percibe el paciente y la unión o vínculo que se crea a nivel psicosocial de catastrofismo, que da lugar a una pobre mejoría en la función. Por otro lado, se entiende como "función", a un término más amplio que incluye dimensiones de movimiento, fuerza, estabilidad articular, dolor e incluso apariencia estética. El término función no es generalmente admitido y sus dimensiones no suelen ser valoradas cuando hablamos de resultados según la valoración del paciente. Discapacidad y función pueden estar relacionados, pero no significa que uno sea igual al otro. (Badalamente 2013)

El DASH, está constituido por un cuestionario de 30 ítems: 21 ítems relacionados con la función física, 6 ítems relacionados con síntomas, y 3 ítems que abordan las limitaciones sociales y la funcionalidad. Por otro lado, cuenta con dos módulos opcionales de 4 ítems: uno dirigido a deportistas y músicos; y otro a la población trabajadora, con el fin de conocer la dificultad o discapacidad para realizar las actividades cotidianas específicas en estas poblaciones.

El instrumento fue diseñado para ser sensible en la detección de discapacidad ocasionada por patologías que afecten cualquier parte de la extremidad superior. Sus primeros usos se centraron en problemas que afectaban las grandes articulaciones proximales como hombro y codo. A nivel de la mano, se ha empleado para valorar

resultados en el Síndrome del Túnel Carpiano (Gay 2003), (Atroshi 2002) (Rosales 2002), en fracturas distales de radio (McDermid 2000) y recientemente para valorar la eficacia de determinadas artrodesis segmentarias del carpo como la fusión de las cuatro esquinas en comparación con la artrodesis total de muñeca (Kluge 2000), y en comparación con la carpectomía total (Cohen 2001).

Se ha demostrado que el DASH es útil para describir individuos con enfermedades o patologías del miembro superior, estimar el pronóstico de dichos sujetos, y para evaluar los cambios clínicos tras tratamiento, en base a estudios previos desarrollados para confirmar su consistencia interna, reproductibilidad, validez y la sensibilidad para detectar cambios de importancia clínica.

El cálculo de la puntuación utilizando el cuestionario DASH es relativamente complicado. Dado que es un instrumento que mide la discapacidad, su rango va desde 0 que equivale a no discapacidad a 100 que significa el peor estado de salud. Es importante entender este sentido de la puntuación ya que los instrumentos que miden el estado de salud van de 0 que equivale a una mala salud a 100 que equivale al mejor estado de salud. Para poder calcular la puntuación es necesario que se hayan contestado al menos 27 de los 30 ítems. La puntuación final se obtiene calculando la media aritmética de las preguntas contestadas, restando 1 y multiplicando por 25: puntuación DASH = (puntuación cuestiones contestadas/número de cuestiones contestadas) -1) x 25. La puntuación Función/Síntomas da "la puntuación cruda". El máximo de puntuación que se puede obtener es 150 y la mínima es 30 (rango = 120). La puntuación cruda es entonces transformada a una escala de 0 a 100; donde cero refleja la ausencia de discapacidad (buena función) y 100 muestra la mayor discapacidad (mala función), por lo tanto, mayor puntuación indica mayor discapacidad.

Los módulos adicionales opcionales deporte/música o trabajo, se calculan de igual forma, no pudiéndose calcular la puntuación si no se han contestado todas las cuestiones de cada uno de estos dos módulos. El objetivo fue incluir ítems en el DASH que pudieran capturar las delicadas dificultades que los profesionales de la música y el deporte o trabajadores pueden encontrar en su ocupación habitual y que a lo mejor no afectan a las actividades cotidianas de la vida diaria. La puntuación cruda máxima que se puede obtener es 20 con un mínimo de 4. El rango de puntuación es 16. La puntuación también es transformada a una escala de 0 a 100, en la que la puntuación más baja implica discapacidad mínima y puntuaciones altas reflejan mayor discapacidad.

Para el manejo de ítems no contestados: si menos del 10% de los ítems (3 preguntas función/síntomas) han sido dejados en blanco por el sujeto que responde el cuestionario, la media de la puntuación del resto de los ítems puede ser sustituido en el lugar de los ítems en blanco. De hecho, si una persona responde sólo 28 ítems con una media de puntuación de "3" y deja en blanco 2 ítems, el valor perdido (missing value) puede ser reemplazado por el valor "3" como media de las otras respuestas. Si más del 10% de los ítems son dejados en blanco, entonces no es posible calcular la puntuación final del DASH. Por la misma regla, ítems en blanco no son aceptados en las escalas o módulos opcionales (sólo 4 ítems).

Como era de esperar, la sensibilidad al cambio del DASH es mayor que la de cuestionarios de salud generales, como el SF-36, y menor que la de cuestionarios específicos que valoran condiciones o regiones específicas asociadas a patologías en concreto, como el instrumento CTS. Entonces, ¿cuándo usamos el DASH?. El DASH es un instrumento excelente para el estudio general del miembro superior, ya que puede ser administrado de rutina sin especificar cuál es la patología en concreto, y lo ideal

sería entregarlo incluso antes de que el paciente fuera valorado por el especialista de miembro superior.

En un estudio realizado sobre pacientes afectados de síndrome del túnel carpiano aplicando el cuestionario específico CTS y el DASH, y subdividiendo este en dos subescalas, una con ítems similares a los ítems del CTS y otra con el resto, se concluyó que el uso combinado de cuestionarios era capaz de detectar detrimento en las subescalas no específicas que serían atribuibles a otros problemas no relacionados con el síndrome del túnel carpiano, sin afectar a la confiabilidad. En este sentido el uso combinado de instrumentos es una herramienta muy útil para definir poblaciones de estudio y para estudios del estado de salud en la población. (Reboso 1998)

El instrumento DASH está disponible en 27 lenguas

(<http://www.dash.iwh.on.ca/translate.htm>) , existiendo la validación al castellano.

(Anexo F)

#### **1.3.4.2.1 Versión Española del Instrumento DASH. Adaptación Cultural y Lingüística**

La versión española fue publicada en el año 2002 en el Journal of Hand Surgery ("Evaluation of the Spanish Version of the DASH and Carpal Tunnel Syndrome Health-Related Quality-of-Life Instruments: Cross-cultural Adaptation Process and Reliability". (R.S.Rosales 2002)

Un instrumento para medir salud debe ser entendido como cualquier otro instrumento de medición, con la diferencia de que vamos a medir una dimensión de la salud como la funcionalidad física, salud mental, función social, etc. Estos instrumentos son cuestionarios desarrollados en países como Canadá, Reino Unido o EEUU. Debido

a ello, los investigadores y cirujanos de habla no anglosajona se enfrentaban a uno de los errores sistemáticos más importantes, ya que solían ser traducidos por el propio investigador para ser aplicados de inmediato en una población estudio, sin tener en cuenta los sesgos socioculturales de la traducción. Para evitar este importante error, el primer paso metodológico en el uso de instrumentos de salud es la adaptación.

El objetivo último y primordial del "Movimiento Outcomes" es el poder comparar resultados a través de diferentes países o comunidades socio-culturales, y para ello es necesario obtener versiones conceptualmente equivalentes con la versión original del instrumento de salud.

Existen diferentes métodos de adaptación socio cultural y lingüístico al respecto. El método utilizado de adaptación fue el de traducción-retrotraducción (Ware 1996) (Alonso 1998) empleado por ejemplo en el proyecto de la IQOLA (International Quality of Life Assessment) para desarrollar diferentes versiones del cuestionario genérico SF-36. Reuniones de traductores, investigadores y pacientes dan lugar a sucesivas versiones, a la vez que los traductores cuantifican la equivalencia conceptual y la dificultad de la traducción. El método traducción- retrotraducción, en el caso de la versión española del DASH, se basó en traducir los cuestionarios originales al castellano por dos traductores bilingües cuya lengua original es el castellano y con experiencia clínica. Las dos nuevas versiones en castellano fueron discutidas simultáneamente entre traductores y uno de los investigadores del proyecto. Otros dos traductores bilingües valoraron la equivalencia conceptual, la claridad de cada frase y cada una de las opciones de respuesta en estas dos primeras versiones del castellano. Una segunda reunión entre traductores e investigadores fue llevada a cabo para obtener una primera versión en castellano de los cuestionarios (versión 1.0). Para valorar la equivalencia conceptual de esta primera versión consensuada en castellano, la versión

1.0 fue retrotraducida al inglés por dos traductores bilingües cuya lengua nativa era el inglés y estaban viviendo en España. Las dos nuevas versiones retrotraducidas al inglés fueron comparadas con la versión original de los cuestionarios para tratar de identificar aquellos ítems o palabras que no presentaban una completa equivalencia conceptual. Un estudio piloto fue realizado en diferentes poblaciones de pacientes para valorar la comprensión de las nuevas versiones 1.0 en castellano y a continuación se obtuvo la definitiva versión 2.0 de los instrumentos en una nueva reunión entre traductores y el equipo de investigación. Finalmente, se procedió a realizar el análisis de la ordinalidad de las opciones de respuesta que se basó en la escala visual o ejercicio de Thurstone. Un instrumento de salud basa sus mediciones en las posibles opciones de respuestas para cada ítem. Así, para un ítem que pregunta sobre la dificultad que presenta para realizar una determinada actividad, las opciones de respuesta pueden ser: "ninguna- leve- moderada- grave- imposible de realizar", teniendo un valor numérico que en este caso va desde 1 (ninguna) hasta 5 (imposible de realizar). Debido a que la percepción del paciente sobre la dificultad que presenta para realizar dicha actividad está influenciada por factores socioculturales era necesario conocer si por ejemplo cuando la población española gradúa una dificultad como "leve", ésta es coincidente con el concepto "leve" que percibe la población donde se desarrolló el cuestionario original, para ello, a los participantes en el estudio se les requería marcar en una línea de 10 cm de longitud (escala visual) la posición relativa de las diferentes opciones de respuesta en relación a las dos opciones extremas, y las opciones intermedias, siendo comparada con las puntuaciones teóricas del cuestionario original. Así el estudio reveló que el orden de las medias de las puntuaciones obtenidas en el ejercicio era coincidente con el establecido en el cuestionario original. Este estudio sirvió para decidir en aquellas opciones de respuesta que tenían más de una traducción posible y en las que sólo era posible incluir

una opción traducida en el cuestionario final en castellano. De esta manera para la opción "severe difficulty" fue incluida la opción traducida "muchacha" en vez de la opción "bastante".

Una vez finalizada la fase de comprobación de las propiedades psicométricas (confiabilidad, validez y respuesta), la versión española del DASH fue valorada por el Comité de Medicina Basada en la Evidencia de la AAOS y del IWH, recibiendo la aprobación como versión oficial en español. (Rosales 2002)

La validez de construcción del módulo específico del cuestionario DASH para medir discapacidad específica de Miembros Superiores fue analizada. Un total de 359 músicos fueron incluidos en este estudio, un 43.5% hombres y un 56.5% mujeres. Las versiones oficiales españolas de los cuestionarios de salud SF-36 y del DASH fueron autoadministrados a la población diana. En orden a demostrar la validez de construcción de este módulo del cuestionario DASH, el equipo investigador estableció la siguiente hipótesis: El grado de dificultad para tocar un instrumento se correlacionará de forma positiva con el grado de discapacidad en los MMSS y de forma negativa con el nivel de calidad de vida (SF-36). Las conclusiones fueron que existe una correlación positiva entre la dificultad para tocar un instrumento musical y la discapacidad en MMSS ( $p= 0.000$ ,  $R= 0.5640$ ) y que el grado de dificultad para tocar un instrumento se correlaciona negativamente con la Calidad de Vida, a mayor dificultad menor calidad de vida, ( $p= 0.000$ ,  $R= -0.240$  a  $-0.3398$ ) influyendo en ello la discapacidad observada en sus MMSS. Todo ello constituyen evidencias científicas sobre la validez de construcción del módulo específico para músicos del cuestionario DASH. (Rosales 2005)

### 1.3.4.3 Cuestionario QuickDASH

Se trata de la versión abreviada del cuestionario DASH que permite una valoración más rápida del resultado. Este cuestionario denominado Quick-DASH, consta de 11 ítems.

Beaton et al. en el 2005, desarrollaron un estudio mediante el cual calcularon que para que cualquier versión del DASH siguiera teniendo un coeficiente de Cronbach alfa de 0.90-0.95, el mínimo número de ítems del cuestionario debía de ser de 11, método denominado "retención conceptual".

Los llamados instrumentos "cortos" tienen la ventaja de ofrecer la misma calidad de información con menos carga para el paciente a la hora de completarlo y una mayor facilidad a la hora de puntuarlo para el médico o investigador, facilitando así su uso y disminuyendo la cantidad de datos perdidos.

El Quick-DASH, con 11 ítems, aborda síntomas y función física en pacientes con uno ó múltiples problemas en la extremidad superior. Los módulos opcionales no cambian del DASH original. Su puntuación es similar a la del DASH, con un rango de 0 a 100, indicando 100 la mayor discapacidad. Sólo es tolerable la pérdida de un ítem (10%).

La alta correlación entre el DASH y el QuickDASH permite comparar ambas puntuaciones, pero no de manera exacta, es decir, 45 puntos del DASH no equivalen a 45 puntos del QuickDASH. (Beaton 2005)

Así como para las diferentes versiones del DASH existen innumerables publicaciones que aportan evidencias científicas sobre la validez, fiabilidad y respuesta de este cuestionario, para la versión reducida "QuickDASH", sólo existen dos publicaciones, una sobre la versión americana y otra sobre la versión sueca, que aportan

datos y evidencias sobre validez y que fundamentan el principio de que la versión reducida del cuestionario es capaz de medir lo que el DASH original está midiendo. Sin embargo, Gummesson y colaboradores exponen en su trabajo que poco se conoce de cómo la puntuación del QuickDASH puede ser interpretada en comparación al DASH. (Gummesson 2006)

En el trabajo, "Validez de Construcción de la Versión Española del Instrumento QuickDASH para medir resultados en Cirugía de la Mano y su valor predictivo en determinar discapacidad de MMSS", (R.S. Rosales, Y. Martín, 2007) se analizó la validez de construcción de la versión española del QuickDASH. Los resultados demostraron que el QuickDASH y el DASH se correlacionan con el instrumento específico del Síndrome del Túnel Carpiano de una manera similar, lo que constituye una prueba de validez de construcción. Iguales resultados han sido publicados con las versiones americanas y suecas; sin embargo, la relación entre la puntuación de la versión completa del DASH no pudo ser establecida por estos autores. La observación de los diagramas de dispersión demuestra que la relación entre el QuickDASH y el DASH es muy similar entre la versión sueca y la española. Pero además, lo que constituyó una innovación del mencionado trabajo es que el análisis de regresión lineal demostró un coeficiente de determinación "R cuadrado" de más de un 90% y con un nivel de residuos mínimo, lo que significa que la proporción de la varianza en la puntuación del DASH puede ser explicado en más de un 90% mediante la variable "QuickDASH", por lo tanto, la versión española del QuickDASH presenta una buena validez de construcción, pudiendo ser utilizado en vez del DASH en la valoración de resultados de pacientes .

## 1.3.4.4 Instrumento Específico del Síndrome del Túnel Carpiano

El síndrome del túnel carpiano o neuropatía por compresión del nervio mediano en la muñeca, es causa de dolor, entumecimiento y calambres en la extremidad superior. Su prevalencia se estima en el 0.1 % de la población general y hasta en un 1-5% de trabajadores. La liberación del túnel carpiano, es una de las cirugías más comunes realizada en la mano, llegando en Estados Unidos a 200.000 casos al año, lo que supone un gran coste económico. La tasa de éxito publicada oscila entre el 70-90 % valorando únicamente parámetros clínicos como la fuerza y otros hallazgos físicos, sin tener en cuenta si el paciente seguía con síntomas o alteraciones en su función. (Levine 1993)

Aún siendo la mononeuropatía periférica más frecuente de la extremidad superior, hay pocos estudios de su prevalencia en la población general, en cambio, sí existen estudios que avalan su relación con diferentes factores de riesgo como la diabetes, el sobrepeso, el embarazo y ciertas actividades laborales. Uno de los estudios más importantes fue el desarrollado por Isam Atroshi en 1999, para estimar la prevalencia del STC en la población general de Suecia, destacando que 1 de cada 5 personas con síntomas (calambres, dolor, entumecimiento) eran objeto de padecer atrapamiento del nervio mediano a nivel del túnel carpiano tanto por la exploración clínica como por el estudio electroneuromiográfico (ENMG). Como muchos síndromes clínicos, la definición de STC difiere de un investigador a otro. Aunque las anomalías en el estudio ENMG han sido consideradas como primordiales para algunos, otros creen que ni son necesarias ni suficientes para el diagnóstico de STC, publicando casos de STC en ausencia de cambios en el ENMG. La mayoría de los autores están de acuerdo que debe existir una concordancia entre los signos clínicos y las anomalías del ENMG.

La variación en la prevalencia del STC se estima que en parte es causada por los diferentes criterios diagnósticos en su estudio. Un diagnóstico erróneo se debe probablemente al uso inapropiado del ENMG que a su vez da lugar a un tratamiento fallido del STC. Aunque el ENMG se ha considerado como el test gold estándar, esta afirmación es cuestionada por 3 razones: primero porque el ENMG resulta anormal cuando la compresión es suficientemente severa para causar anomalías estructurales en el nervio mediano, y la conducción nerviosa puede ser normal a pesar de la presencia de datos clínicos importantes. Estos serían los casos que mejoran tras cirugía a pesar del ENMG. Segundo, la velocidad de conducción nerviosa se define como anormal cuando tiene valores más bajos de 2.5 % - 5 % de lo normal; sin embargo la velocidad en pacientes asintomáticos también puede dar resultados muy bajos, dando resultados positivos en pacientes sanos. Y tercero, el punto de corte para definir anormalidad en la medición de la velocidad varía ampliamente. Algunos especialistas en cirugía de la mano no usan el resultado del ENMG como criterio diagnóstico de STC, sino la exploración física y clínica. La importancia clínica que se atribuye a los hallazgos en la exploración varía de un experto a otro. Con la finalidad de encontrar unos criterios clínicos estandarizados para establecer un criterio diagnóstico para el STC, B. Graham et al. realizó un estudio basándose en un proceso metodológico amplio desarrollado en 3 estadios (B.Graham 2006). Los 6 criterios clínicos que contribuyeron significativamente ( $p < 0.05$ ) al diagnóstico de STC fueron: entumecimiento en el territorio del mediano, entumecimiento nocturno, debilidad/atrofia tenar, signo de túnel positivo, y pérdida de la capacidad de discriminación.

Levine en 1993, desarrolló un instrumento específico para la valoración de la severidad de síntomas y el estado funcional del paciente con síndrome del túnel carpiano, de probada validez, fiabilidad y sensible a los cambios clínicos. Para su

desarrollo, fueron consultados cirujanos de mano, reumatólogos y pacientes, para determinar 6 síntomas claves : dolor, parestesias, entumecimiento, debilidad, síntomas nocturnos y sobre todo el estado funcional.

El Instrumento CTS es un cuestionario que consiste en 2 escalas. Una escala que evalúa la severidad de los síntomas con 11 ítems que valoran la severidad, frecuencia, y duración de los síntomas (durante el día, durante la noche, entumecimiento, calambres, debilidad, dificultad en coger pequeños objetos) con múltiples respuestas que van desde 1 punto (más leve) a 5 puntos (más severa). La puntuación global para la escala de severidad de síntomas se calcula con la media de los 11 ítems. En la escala para valoración del estado funcional con 8 ítems, se identificaron actividades comunes que se ven afectadas en pacientes con STC.

En un principio fueron 12 ítems, pero después de un estudio en el que la amplia mayoría dejaba sin responder 4 ítems (conducir, escribir a máquina, hacer deporte, tocar instrumentos), se creó la escala con 8 ítems (escribir, abrocharse la ropa, mantener un libro mientras se lee, mantener un teléfono, abrir un tarro, hacer las tareas del hogar, cargar bolsas de la compra, bañarse-vestirse). Estas actividades son comunes en un amplio espectro de pacientes, tanto en gente trabajadora joven como en ancianos. Este abordaje disminuye la pérdida de datos que puede comprometer al resultado del estudio.

En algunas circunstancias, la escala puede ser modificada añadiendo ítems específicos según la población que estudiemos, por ejemplo si se trata de una muestra sólo con pacientes trabajadores.

Las respuestas fueron puntuadas desde 1 punto (sin dificultad) a 5 puntos (imposible de realizar). La puntuación global para el estado funcional es el resultado de la media de los 8 ítems. Los ítems que son dejados sin contestar no son incluidos en la media.

La reproductibilidad, la consistencia interna, la validez, y la sensibilidad para detectar cambios clínicos fueron evaluados en un estudio clínico que fue aprobado por el Human Investigations Committee of Brigham and Women's Hospital. La confiabilidad refleja si el mismo resultado es obtenido en varias administraciones asumiendo que no hay cambios clínicos. Se obtuvo una alta confiabilidad con un coeficiente de correlación de Pearson ( $r$ ) en la escala de severidad de síntomas de 0.91 y en la escala funcional de 0.93. Un coeficiente de correlación ( $r$ ) de 0 indica no correlación y de 1.0 indica una perfecta reproductibilidad. La consistencia interna mostró un coeficiente alfa de Cronbach de 0.89 para la escala de severidad de síntomas y de 0.91 para la escala de estado funcional, esto implica que los ítems dentro de cada escala están relacionados entre sí, no son independientes. En término medio, se considera que un coeficiente alfa de Cronbach de 0.8 es considerado bueno, y un valor de 0.9 es considerado excelente. La sensibilidad al cambio clínico fue determinada mediante el tamaño del efecto, calculado como la media de la diferencia entre la puntuación preoperatoria y postoperatoria dividido entre la desviación estándar de la diferencia. El resultado del tamaño del efecto en la escala de severidad de síntomas fue de 1.4 y en la escala del estado funcional de 0.82. En general, un efecto del tamaño de más de 0.5 es considerado un resultado moderado y más de 0.8 un resultado importante. (Levine 1993)

Existe un amplio consenso que defiende la necesidad de establecer valoraciones rigurosas basadas en instrumentos de probada validez, fiabilidad y sensibilidad para distinguir intervenciones que son efectivas de aquellas que no lo son y valorar el beneficio del tratamiento quirúrgico versus conservador del síndrome del túnel carpiano. Es aceptado que el síndrome del túnel carpiano severo que presenta atrofia tenar y/o disminución severa de la sensibilidad posiblemente requiera cirugía y que

pacientes con síntomas y limitaciones moderadas sean tratados en principio de manera conservadora, pero la decisión es difícil. Existen estudios de alta evidencia científica, ECA y estudios de cohorte, que apoyan la efectividad del tratamiento quirúrgico en casos seleccionados (liberación abierta o endoscópica) (Atroshi 2006). En la revista médica Británica *The Lancet* (2009), se publicaron los resultados de un estudio multicéntrico de ECA en el que se comparaba el tratamiento quirúrgico con el conservador (Jarvik 2009)

En el año 2009 Atroshi desarrolló una versión reducida del Instrumento CTS, mediante un análisis metodológico de las escalas severidad de síntomas y estado funcional, resultando en la eliminación de 4 ítems y fusionando otros 2 ítems en la escala de valoración de severidad de síntomas. De este modo, se desarrolló una escala con 6 ítems ([www.jhandsurg.org](http://www.jhandsurg.org)), que no sólo es más corta por el menor número de ítems, sino por su diseño (los 6 ítems quedan dispuestos en la misma horizontal). Asimismo, quedó demostrada su buena consistencia interna, test retest reliability, y validez comparado con la escala de 11 ítems. Además, Atroshi et al, desarrollaron una escala de 2 ítems para valorar el dolor palmar en la mano después de la liberación del túnel carpiano. Después de la liberación del nervio mediano a nivel del túnel carpiano, es muy común el dolor en la cicatriz y el dolor proximal, que varía en severidad y duración. Diferentes técnicas tanto endoscópicas o abiertas, mediante mini-incisiones se han desarrollado con el objetivo de disminuir la morbilidad de esta intervención, pero el dolor palmar no había sido estudiado mediante ningún Instrumento de valoración de resultados. Con el desarrollo de la escala 2-item dolor palmar ([www.jhandsurg.org](http://www.jhandsurg.org)) se puede medir de manera específica el dolor palmar relacionado con la cirugía. Esta escala consiste en 2 ítems que preguntan sobre la severidad del dolor debido a la cicatriz en la palma de la mano (5 respuestas posibles) y sobre el

grado de limitación en la mano debido a ese dolor (6 respuestas posibles). La escala es puntuada en un rango que va desde 0 (no dolor palmar o limitación de la actividad) a 100 (el dolor más severo). La puntuación del 6-CTS se realiza de forma similar al 11-CTS; cada ítem es puntuado en una escala que va desde 1 (no síntomas) a 5 (muy severo). Para cada paciente, se calcula la media de los ítems contestados, sólo se permite un ítem no contestado, que nos dará un resultado que va desde 1 (mejor) a 5 (peor).

Hay que tener presente que los instrumentos específicos, en este caso el Instrumento de Levine para STC, no es un instrumento diagnóstico, ya que los ítems que miden las escalas de síntomas e incapacidad funcional pueden ser causados por otras neuropatías compresivas u otras causas patológicas de la extremidad superior. Teóricamente, los pacientes con otras neuropatías compresivas pueden tener respuestas similares a los pacientes con STC. El diagnóstico de STC debe ser hecho independientemente al instrumento STC. Un examen físico y clínico es esencial. En 1997, Atroshi valoró a todos los pacientes mediante el diagrama de sensibilidad de la mano de Katz and Stirrat que tiene una alta sensibilidad y especificidad para el diagnóstico de STC. La probabilidad de tener STC es relacionado con la distribución de síntomas que el paciente marca en el diagrama. Según el diagrama de Katz- Stirrat se puede clasificar al paciente en 4 categorías: clásico, probable, posible o improbable STC. De este modo, los pacientes con un clásico y probable STC, tenían una puntuación más alta en la escala de severidad de síntomas que los que tenían un posible o improbable STC según el diagrama. En cambio, no había diferencias en la escala funcional. De este modo, se demostró que la escala de severidad de síntomas era más específica para valorar resultados en pacientes con STC que con otras neuropatías compresivas, por lo que el instrumento STC es más específico para la compresión del

nervio mediano. De manera inversa, la escala de valoración funcional no muestra diferencias con tener STC u otra neuropatía. (Atroshi 1997)

Valorar la satisfacción del paciente no es fácil. Los síntomas pueden estar relacionados por otros factores como la propia expectativa del paciente, su personalidad e incluso el trato con el cirujano. En muchos estudios sobre resultados en STC , el énfasis se ha centrado en los síntomas y signos provocados por la compresión nerviosa y el alivio con la cirugía, mientras que a la mejoría funcional no se le ha prestado mucha atención. La identificación de factores que se correlacionen con la satisfacción del paciente es primordial (Atroshi 1998). Se identificaron 11 actividades de la vida diaria que pudieran ser afectadas por padecer STC, tales como, escribir, abrocharse botones, la higiene personal, vestirse, peinarse, abrir tarros, partir rebanadas de pan, usar el tenedor y cuchillo, usar una llave, conducir y participar en actividades de ocio. Para cada actividad, la respuesta fue puntuada como 1 (no dificultad), 2 (dificultad media), 3 (dificultad moderada), 4 (dificultad severa) y 5 (imposible de realizar la actividad debido a los síntomas de la mano). Una puntuación mayor indica una peor función. Una mayor puntuación preoperatoria y una valoración precoz de resultados se ha relacionado con un nivel más bajo de satisfacción. A largo plazo esta satisfacción podría cambiar. Existe además una probada relación entre la escala de severidad de síntomas y el grado de insatisfacción del paciente, no así con las medidas convencionales del grado de fuerza o sensibilidad, lo que reafirma el uso de instrumentos específicos.

Existen diferentes estudios que comparan la sensibilidad al cambio de instrumentos para determinar cuál es el más apropiado a la hora de realizar nuestro estudio. Uno de los más destacados fue el realizado por Gay y Amadio (Gay 2003) en el que se comparó la sensibilidad al cambio del DASH, instrumento CTS y SF-36 en

pacientes operados del STC. La escala de severidad de síntomas del instrumento CTS es 4 veces más sensible y la escala de estado funcional es 2 veces más sensible que la medición de la fuerza. El instrumento SF-36 es el instrumento genérico más usado, y está demostrada la validez de las subescalas de rol físico y rol dolor corporal para seguimiento de patologías musculoesqueléticas. El DASH es un instrumento específico que permite comparar el impacto en diferentes condiciones musculoesqueléticas. La sensibilidad al cambio es uno de los diferentes factores que deben ser considerados a la hora de elegir un instrumento. No sólo se trata de detectar un cambio, si no, que este cambio sea clínicamente importante (relevancia). Pequeños cambios pueden ser clínicamente importantes y grandes cambios pueden ser no significativos dependiendo del contexto. La satisfacción del paciente es un importante "medidor" de esta valoración. Como resultado se obtuvo que la sensibilidad al cambio del DASH es lo suficientemente grande (indicado por el ES, SRM y correlación con el instrumento STC) y que puede ser usado como instrumento de valoración primario tras cirugía del síndrome del túnel carpiano cuando la duración del seguimiento es superior a las 12 semanas. Aunque con el DASH se necesitará mayor tamaño de la muestra que con el STC, el DASH será capaz de comparar resultados del impacto del STC con otras patologías del miembro superior.(Gay 2003)



## 1.4 Justificación y Estructura de la Tesis

Como se ha comentado anteriormente, los instrumentos PRO (Patient Reported Outcomes) y la valoración de los niveles de evidencia en cirugía de la mano son esenciales en la práctica médica. Los avances tecnológicos han permitido que la Cirugía de la Mano alcance límites insospechados, pero este desarrollo técnico no se ha acompañado de una mejora en la calidad de los trabajos de investigación clínica publicados. El nivel de los Ensayos Clínicos y estudios observacionales prospectivos, publicados en las principales revistas científicas relacionadas con la mano y miembros superiores (MMSS), es muy bajo y de muy poca calidad. (Amadio 1996) (Gummenson 2004)

Gummenson et al en el 2004, realizó la primera y única valoración de la calidad de los ECA publicados entre 1999 y 2002 utilizando la escala de Jadad modificada en las cuatro revistas más importantes relacionadas con la patología músculo-esquelética de los MMSS. Conjuntamente, analizó el tipo de variables - resultados utilizadas en dichos ECA llegando a la conclusión que los pocos ECA utilizados presentaban baja calidad.

En la evaluación de la eficacia de una intervención, un aspecto importante es el tipo de medidas de resultado utilizadas y si abarcan aspectos importantes de los efectos del tratamiento. Las medidas estandarizadas en base a la opinión del paciente, denominadas instrumentos PRO, han sido desarrollados para evaluar la salud en relación al nivel de calidad de vida percibida, como por ejemplo, los instrumentos que preguntan acerca de limitaciones en actividades de la vida diaria, o bien, para valorar la salud en base al nivel de participación, como el impacto de la enfermedad en actividades de ocio o sociales. Estos instrumentos que valoran la opinión del paciente

en torno a su calidad de vida, son los que deben hacer mejorar la calidad de nuestra asistencia. Este principio ha constituido una de las más importantes fuerzas externas que han impulsado la entrada a una nueva era de la evaluación médica basada en la opinión del usuario del servicio sanitario que ha sido denominada por el editor jefe del New England Journal of Medicine como Movimiento Outcomes "Outcomes Movement". El objetivo último y primordial del "Movimiento Outcomes" es el poder comparar resultados a través de diferentes países o comunidades socio-culturales, y para ello es necesario obtener versiones conceptualmente equivalentes con la versión original del instrumento de salud. (Rosales RS 2002), (Atroshi 2002) (Atroshi 2000). Una vez obtenidas las versiones estándar en español de los instrumentos PRO, se deben probar sus propiedades psicométricas: consistencia interna, confiabilidad, validez de construcción y respuesta o sensibilidad para detectar cambios de importancia clínica. Recientemente, se han desarrollado las versiones cortas de los mismos que son el QuickDASH y el 6-ítems CTS (Rosales 2013), (Atroshi 2009), (Atroshi 2011) teniendo versiones adaptadas estándar en español, pero no evidencia sobre sus propiedades psicométricas.

En general, para llevar a cabo un estudio de investigación clínico usando los instrumentos PRO debemos tener en consideración lo siguiente: 1) antes de iniciar cualquier estudio se debe obtener el permiso del Comité Ético Investigador de nuestra institución, 2) establecer el tamaño de la muestra para evitar aceptar como correcta la hipótesis nula (fallo en identificar una diferencia estadísticamente significativa cuando existe una diferencia verdadera) y prevenir sesgos innecesarios. En general, el número de sujetos de la muestra debe ser de un 10% a un 20% superior al tamaño de la muestra calculado para tener en cuenta el número de pacientes que no terminan el seguimiento. 3) Tiempo de seguimiento; debe ser suficiente pero no excesivo. En general, el estudio

no debe finalizar antes de llegar al tiempo "plateau" tras el tratamiento. Por ejemplo, Rozenthal demostró que la movilidad de la muñeca tras una fractura de radio mejoraba hasta los 12 meses tras la fijación de la misma. 4) manejo de los datos; siguiendo las normas del Comité Científico sobre la privacidad de datos del paciente. 5) Pérdida de datos: en casi todos los estudios clínicos hay pérdida de datos. La cantidad de datos perdidos puede afectar a la validez de una medida. No hay consenso sobre la cantidad de datos que se tienen que perder para que la valoración sea inaceptable. De hecho, cada instrumento de valoración de resultados, proporciona los pasos a seguir cuando hay pérdida de ítems y calcular la puntuación final al respecto. Por ejemplo, en el Instrumento The Michigan Hand Outcomes Questionnaire (MHQ) sólo es válido si se contestan más del 50% de sus ítems. Para los ensayos clínicos aleatorizados se establece que debe existir más de un 80% de seguimiento.

En la adaptación cultural y lingüística de un instrumento se deben analizar sus propiedades psicométricas: confiabilidad, validez y sensibilidad al cambio. Es un largo proceso que no acaba con la traducción- retrotraducción. En la adaptación de la versión española del DASH se ha analizado la confiabilidad en pacientes con síndrome del túnel carpiano, sin embargo, la sensibilidad al cambio y la validez de construcción, todavía no habían sido evaluadas. Ha sido demostrado que el DASH es comparable con el cuestionario del túnel carpiano (CTS) para valoración en muñeca y mano en general (Beaton 2001) y también que el DASH es lo suficientemente sensible para su uso en estudios del síndrome del túnel carpiano y que su sensibilidad para detectar cambios de importancia clínica (a las 12 semanas o más después del tratamiento) es intermedio en comparación con el cuestionario específico del túnel carpiano (Levine 1993) y con el instrumento genérico SF-36. )

El cuestionario CTS desarrollado por Levine et al. (Levine 1993) ha sido uno de los instrumentos de valoración (PRO) más usados en las dos últimas décadas. El cuestionario CTS original consiste en 2 escalas: severidad de síntomas (SS) con 11 ítems y estado funcional. Usando el factor análisis y la metodología de la teoría de la respuesta de los ítems, Atroshi et al. desarrollaron una versión reducida de la escala de severidad de síntomas con sólo 6 ítems, con el objetivo de reducir la carga al paciente sin perder las propiedades psicométricas del cuestionario. Ellos demostraron que la nueva versión reducida tenía un buen nivel de confiabilidad, validez y sensibilidad.(Atroshi 2009).

La adaptación transcultural al español del DASH y CTS original de 11 ítems fue realizada por nuestro equipo de trabajo y publicada en el 2002 (Rosales 2002), cuya línea de investigación actual se centra en el desarrollo de las versiones reducidas de los mismos.

El uso de instrumentos de salud específicos para medir la valoración de resultados está en crecimiento. La cirugía del síndrome del túnel carpiano es una de las más frecuentes realizadas hoy en día en un servicio de cirugía de mano, de ahí la importancia de valorar sus resultados.

Una vez revisado el estado actual de la medicina basada en la evidencia, el movimiento outcomes y del uso de instrumentos PRO en la valoración de resultados, procedemos a plantear los objetivos de la presente Tesis Doctoral.

El planteamiento de nuestros objetivos los centramos en tres hipótesis:

## 1. Hipótesis Primera

Estudiar el nivel actual de evidencia científica de la investigación clínica en cirugía de la mano, evaluando el nivel de evidencia de los artículos de investigación clínica publicados en la Revista Iberoamericana de Cirugía de la Mano (RICMA) comparándolo con la Revista Europea de Cirugía de la Mano (JHSE) y la Revista Americana de Cirugía de la Mano (JHSA), con el objetivo de conocer si el nivel de evidencia científica en cirugía de la mano es dependiente del tipo de revista y conocer la calidad de nuestros artículos para establecer un punto de partida para enfocar cambios de mejora en nuestros estudios.

## 2. Hipótesis Segunda

Estudiar el uso de instrumentos PRO específicos de Miembro Superior, en concreto la versión española del DASH analizando los siguientes aspectos:

- Determinar la validez de construcción de la versión española del DASH.
- Estudiar la capacidad de detectar cambios de importancia y significación clínica comparado con las medidas de la exploración física tradicional para medir resultados tras la cirugía del túnel carpiano.
- Estudiar el comportamiento de la versión española del DASH aplicada a pacientes con síndrome del túnel carpiano, frente al instrumento genérico SF-36 y frente al instrumento específico CTS.

### **3. Hipótesis Tercera**

Estudiar la confiabilidad, validez de construcción y la sensibilidad al cambio, de la versión corta española del cuestionario específico 6-CTS (instrumento PRO específico de una patología) para medir resultados tras tratamiento en pacientes afectados de síndrome del túnel carpiano y así determinar su utilidad.

# 2

## **MATERIAL Y MÉTODO**

### **2.1 Hipótesis Primera: Estudio del Nivel de Evidencia en Cirugía de la Mano**

Para determinar el nivel de evidencia científica de los artículos de investigación clínica publicados en las revistas de más relevancia en cirugía de la mano se estableció la siguiente población de estudio y metodología de trabajo.

Para demostrar nuestro objetivo, se estableció como hipótesis nula ( $H_0$ ) que la variable "nivel de evidencia científica" era independiente de la variable "tipo de revista".

## 2.1.1 Población de estudio

Los criterios de inclusión fueron los siguientes: todos los artículos de investigación clínica publicados entre 2005 y 2009 en la Revista Iberoamericana de Cirugía de Mano (RICMA), considerada como la revista y órgano de expresión de la Sociedad Española de Cirugía de Mano y de las que son sociedades Científicas Afiliadas, la Sociedad Portuguesa de Cirugía de Mano de Latino América; la Revista Europea de Cirugía de Mano (JHSE) y la Revista Americana de Cirugía de la Mano (JHSA), como revistas oficiales de la Federación de las Sociedades Europeas de Cirugía de Mano (FESSH) y la Sociedad Americana de Cirugía de Mano (ASSH) respectivamente.

Los criterios de exclusión fueron los siguientes: estudios animales, anatómicos y de cadáver, estudios básicos de ciencias, cursos de instrucción, suplementos de resumen, publicaciones cortas, cartas al editor y artículos de revisión no fueron considerados factibles para el estudio.

Un total de de 932 artículos de investigación clínica siguieron los criterios de inclusión y exclusión (RICMA 60, JHSE 461 y JHSA 411).

## 2.1.2 Evaluación del Nivel de Evidencia

Los artículos que siguieron los criterios de inclusión y exclusión fueron asignados de forma aleatoria a dos observadores independientes (YMH, LRM), con una larga experiencia clínica en cirugía de la mano y bien familiarizados con la Medicina Basada en la Evidencia. Un número igual aproximado de artículos de cada una de las revistas fue evaluado por cada observador. El nivel de evidencia científica para cada

artículo fue evaluado en base a "Niveles de Evidencia Científica - Centro de Oxford para la Medicina Basada en la Evidencia " (Oxford 2009). La valoración fue a ciegas en relación a cualquier información previa del nivel de evidencia de los artículos analizados (ej: "nivel de evidencia y tipo de estudio", el cual es incluido en la JHSA con el resumen de cada artículo desde 2006). Los artículos fueron catalogados de acuerdo a su nivel de evidencia siendo el nivel I (mayor nivel de evidencia, ej: revisiones sistemáticas (RS), meta-análisis (MA), y estudios individuales controlados y aleatorizados con estrechos intervalos de confianza), el nivel II (estudios de cohorte), el nivel III (estudios de caso-control), nivel IV (series de casos y estudios de corte o transversales). y el nivel 5 (menor nivel de evidencia: la opinión de experto y la observación personal). La evaluación de los sub-grupos de nivel de evidencia, así como los tipos de artículos científicos no fueron realizados por los observadores.

### **2.1.3 Análisis de la confiabilidad**

La confiabilidad de la evaluación fue hecha en base al análisis del error intra-observador e inter-observador. Una muestra aleatoria de 30 artículos de investigación clínica, de un total de 872 artículos, publicados en inglés (461 de JHSE y 411 de JHSA), fueron evaluados por dos observadores independientes asignados para el estudio. Después de 15 días, se hizo una segunda evaluación con el orden de los artículos cambiado. Ningún artículo de la RICMA fue incluido en la muestra de estudio para el análisis de la fiabilidad. Esto se hizo para evitar un sesgo de información (Page RM 1995), porque las diferentes lenguas presentes en la RICMA (Español y Portugués), podrían aumentar la fiabilidad intra-observador. La fiabilidad intra e interobservador fue estudiada usando el test Coeficiente Kappa con un nivel de significancia de 0.05.

## 2.1.4 Análisis estadístico

Para la evaluación de los resultados, el número de artículos clasificados para cada nivel de evidencia fue expresado como un porcentaje del total de artículos que cumplieron con los criterios de inclusión y exclusión durante el periodo de tiempo estudiado. Las frecuencias observadas para cada nivel de evidencia de cada revista fueron comparadas con las frecuencias esperadas usando un test chi-cuadrado ( $\chi^2$ ) para variables categóricas con un nivel de significancia de 0.05.

## **2.2 Hipótesis Segunda: Estudio de las propiedades psicométricas de la versión española del DASH en la valoración de resultados de pacientes operados del Síndrome del Túnel Carpiano**

Para estudiar el uso de instrumentos PRO específicos del miembro superior, determinar la validez de construcción de la versión española del DASH, y comparar sus resultados con las medidas estándares de exploración física y con otros instrumentos genéricos (SF-36) y específicos de la patología (CTS)) se seleccionó una población de estudio de pacientes con diagnóstico de síndrome del túnel carpiano y se aplicó el siguiente método.

### **2.2.1 Población estudio**

El protocolo del estudio fue aprobado por el Comité Ético Investigador de nuestra institución. El consentimiento informado se recogió de cada paciente antes de su inclusión.

Un total de 42 pacientes con el diagnóstico de STC bajo criterios clínicos y neurofisiológicos fueron incluidos en este estudio.

Los criterios de inclusión fueron: 1) entumecimiento u hormigueos con o sin dolor en al menos dos de los dedos que corresponden al territorio del nervio mediano (Atroshi 1999; Gay 2003); 2) aumento de los síntomas con los test de provocación del

STC (test de Phalen positivo, y /o Phalen inverso positivo) (Gay 2003); 3) síntomas durante más de 2 meses (Atroshi 1999); 4) fallo con el tratamiento conservador (Atroshi 1999) y 5) estudio electrofisiológico positivo (Amadio 1996; Atroshi 1999; Gay 2003).

Los criterios de exclusión fueron: 1) signos clínicos o electrofisiológicos de compresión nerviosa proximal, 2) diabetes o cualquier enfermedad metabólica, artritis reumatoide u otra patología general inflamatoria (Amadio 1996; Atroshi 1999; Gay 2003).

### **2.2.2 Diseño Clínico**

El diseño clínico empleado fue un estudio de cohorte, con medidas un día antes de la intervención y 12 semanas después de la cirugía. La anestesia empleada fue un bloqueo del plexo braquial y la operación la realizó el mismo cirujano.

### **2.2.3 Instrumentos y Medidas**

Un cirujano de mano experto e independiente, evaluó todos los pacientes el día de antes de la cirugía y 12 semanas después.

El examen físico se basó en medir la destreza manual, los dos puntos de discriminación (2 PPD), la fuerza de puño, la Key Pinch, la pinza pulpejo-pulpejo y el rango de movilidad de la muñeca (ROM). La valoración siguió un protocolo establecido: la destreza se midió a través del pick-up test: se le pidió a los pacientes que cogieran 5 monedas de una mesa, grabando la media de tiempo que tardaban en tres intentos. Los 2 PPD fueron valorados en cada borde radial y cubital del pulpejo de cada

dedo y después medidos sobre una escala de 5 a 15 mm. La media del valor de los tres dedos radiales se usó para el análisis. La fuerza de puño se midió con el dinamómetro de Jamar (Preston, Jackson, MI), la Key Pinch y la pinza pulpejo-pulpejo con el medidor de presión (B&L, Santa Fe Springs, CA), registrando la media de 3 intentos. El rango de movilidad de la muñeca se midió con un goniómetro regular, registrando la movilidad total activa (flexión + extensión + desviación cubital + desviación radial).

Con el fin de demostrar la hipótesis de la validez de construcción, la versión española del instrumento CTS (Rosales 2002) y la versión estándar española del instrumento genérico SF-36 (Alonso 1995) fueron autoadministrados a la población estudio siguiendo el mismo diseño clínico con medidas recogidas el día de antes y 12 semanas después de la cirugía.

## 2.2.4 Análisis de datos

La sensibilidad al cambio para cada medida, fue evaluada calculando el tamaño del efecto (ES=effect size) (media del cambio de antes del tratamiento a después dividido por la desviación estándar (DS) pre-tratamiento) y la media estandarizada de la respuesta (SRM: standardized response mean) (media del cambio dividida por la DS del cambio).

Para el análisis de la validez de construcción, la sensibilidad al cambio del DASH fue comparada con la sensibilidad al cambio de la escala estado funcional y escala de severidad de síntomas del instrumento del túnel carpiano (CTS), y con las dimensiones de las subescalas de rol físico (SF-36rp) y dolor corporal (SF-36bp) del instrumento SF-36.



## **2.3 Hipótesis tercera: Estudio de las propiedades psicométricas de la versión corta española del Instrumento 6-CTS**

Para estudiar la confiabilidad, validez de construcción, y sensibilidad al cambio de la versión corta española del cuestionario específico para medir resultados en el síndrome del túnel carpiano, el 6 ítems CTS (instrumento PRO específico de una patología) se procedió con el siguiente material y método.

### **2.3.1 Población estudio**

El protocolo del estudio fue aprobado por el Comité Ético Investigador de nuestra Institución. Se recogió el consentimiento informado de cada paciente antes de su participación.

La población estudio la constituyeron 40 pacientes (26 mujeres), con una edad media de 54 años (DS: 11), que formaban parte de la lista de espera quirúrgica de liberación del túnel carpiano del Sistema Sanitario Español, en concreto del Servicio Canario de Salud.

Los criterios de inclusión fueron: 1) parestesias o entumecimiento con o sin dolor en al menos 2 de los 4 dedos radiales (Atroshi 1999, Gay R E 2003), 2) aumento de los síntomas con los test de provocación (test de Phalen y/o Phalen inverso positivo, 3) síntomas durante más de 2 meses (Atroshi 1999), 4) fallo en el tratamiento conservador y 5) test de conducción nerviosa positivo de neuropatía del mediano a nivel de la muñeca (latencia distal motora > 4.5 milisegundos, latencia sensitiva muñeca-

dedo >3,5 milisegundos, o velocidad de conducción sensitiva a nivel del túnel carpiano < 40 metros/segundo) (Atroshi 2006, Kimura 1989).

Los criterios de exclusión fueron: 1) signos clínicos o electrofisiológicos de compresión nerviosa proximal, 2) diabetes o cualquier enfermedad metabólica, artritis reumatoide u otra patología general inflamatoria. (Amadio 1996; Atroshi 1999; Gay 2003).

### **2.3.2 Diseño Clínico**

El diseño clínico empleado fue un estudio de cohorte prospectivo con dos mediciones antes de la cirugía separadas una semana, y otra medición 12 semanas después de la cirugía. La anestesia empleada fue un bloqueo del plexo braquial y la operación la realizó el mismo cirujano.

### **2.3.3 Instrumentos y Medidas**

La versión española del 6-item CTS (Rosales 2013), los 11 ítems de valoración de discapacidad del brazo, hombro y mano del cuestionario QuickDASH ([www.iwh.on.ca](http://www.iwh.on.ca)) y el cuestionario Euroqol 5D (EQ-5D) (Badia X 1999), fueron completados por los pacientes en la consulta externa.

El 6-CTS midió la severidad de síntomas relacionados con el túnel carpiano. Está constituido por 6 ítems. 5 de estos 6 ítems, son similares a los correspondientes de la escala de severidad de síntomas de 11 ítems (cuestionario original SS-CTS) y el ítem restante (resultado de la unión de 2 síntomas de la escala de severidad de síntomas) tiene el texto de los 2 ítems.

Sin embargo, el 6 ítems CTS, tiene un diseño completamente diferente y mejorado. La puntuación es similar a la de severidad de síntomas de 11 ítems; para cada paciente la respuesta al ítem es puntuada de 1 (mejor) a 5 (peor) y la media de los 6 ítems nos dará la puntuación del 6-CTS (sólo es permitido la pérdida de 1 ítem).

El QuickDASH es la versión corta del instrumento DASH. Constituido por 11 ítems y que puntúa de 0 (mejor = no discapacidad) a 100 (peor = alta discapacidad). Por lo menos 10 de los 11 ítems deben ser contestados para obtener la puntuación final. Cada ítem se puntúa de 1 a 5, después todos los valores de los 11 ítems son sumados y su media nos da una puntuación de 1 a 5, que es transformada a una puntuación de 0 a 100 restando 1 y multiplicando por 25. Esta transformación se realiza para que sea más fácil de comparar esta puntuación con otras medidas con escalas de 0 a 100.

El instrumento de salud EQ-5D comprende un cuestionario de 5 ítems y una escala analógica visual. Las 5 dimensiones en el EQ-5D son: movilidad, cuidado personal, actividades cotidianas, dolor/malestar y ansiedad/depresión. La puntuación para cada una de las variables es de 1 a 3 puntos, donde 1 = ningún problema, 2 = problema moderado, y 3 = problema extremo. Una vez los datos de cada dimensión son analizados, la puntuación del EQ-5D se genera usando el 5D<sup>TM</sup> Scoring algorithm for Excel (<http://www.ahrq.gov/professionals/cliniciansproviders/resources/rice/EQ5Dscore.html>). El EQ-5D puntúa en un rango que va desde - 0.59 (peor) a 1 (mejor estado de salud). No se podrá puntuar si hay pérdida de respuestas en alguna dimensión.

Para valorar la validez de construcción, la hipótesis establecida fue que la puntuación del instrumento 6-CTS debería tener una correlación de moderada a fuerte con la puntuación del instrumento QuickDASH y de una baja a moderada negativa correlación con el índice EQ-5D.

Para determinar la fiabilidad con el test-retest se administró al paciente por segunda vez el instrumento 6-CTS una semana después de la primera.

## 2.3.4 Análisis de Datos

La consistencia interna de la confiabilidad fue valorada mediante el coeficiente alfa de Cronbach (alfa  $>0.7$ , indica una buena consistencia interna). El test-retest de la confiabilidad fue valorado con el coeficiente de correlación intraclass (ICC) y con la comparación de la media de las 2 puntuaciones consecutivas del 6-CTS con el t-test. Para el análisis de la diferencia media para el test retest se necesitó una muestra de 19 pacientes para detectar una diferencia clínica importante de 0.9 en la puntuación 6-CTS, asumiendo una SD de 0.7 (Atroshi 2009), prueba de las dos colas, poder del 80% y un nivel de significancia del 0.05.

La hipótesis de la validez de construcción fue analizada con el coeficiente de correlación de Pearson ( $r$ ), usando 0.01 como nivel de significancia.

La sensibilidad al cambio para cada medida, fue evaluada calculando el tamaño del efecto (ES=effect size) (media del cambio de antes del tratamiento a después dividido por la desviación estándar (DS) pre-tratamiento) y la media estandarizada de la respuesta (SRM: standarized response mean) (media del cambio dividida por la DS del cambio). Se considera una mejoría clínica significativa cuando el  $ES > 0.8$ .

# 3

## RESULTADOS

### 3.1. Hipótesis Primera: Estudio del Nivel de Evidencia en Cirugía de la Mano

El análisis inter-observador mostró un Kappa de 0.617, con un error asintótico estándar (SE) de 0.117. El análisis intra-observador presentó un Kappa de 0.66 (SE 0.114) para el observador 1, y un Kappa de 0.751 (SE 0.103) para el observador 2. Todos los valores Kappa fueron significativos ( $p < 0.001$ ).

Más del 80% de los artículos publicados en RICMA y la JHSE; y un 67.6% en la JHSA presentaron un nivel 4 de evidencia. Ningún artículo con nivel I y II de evidencia fue publicado en la RICMA comparado con la JHSE (0.9% de nivel I y 5.0%

de nivel II) y la JHSA (8.3% de nivel I y 10 % de nivel II). El porcentaje de artículos con nivel III, publicados en la RICMA (16.7%) fue mayor comparado con la JHSE (11,1%) y la JHSA (14,1%). (Tabla)

Todos los resultados fueron estadísticamente significativos ( $\chi^2=63.945$ ;  $p<0.001$ ) y la Hipótesis nula ( $H_0$ ) fue rechazada.

Los resultados han demostrado con un nivel de fiabilidad bueno-excelente que la variable "nivel de evidencia científica" es dependiente de la variable "tipo de revista". Asimismo se ha demostrado que el acuerdo inter- observador e intra-observador entre dos cirujanos con experiencia clínica en cirugía de la mano y familiarizados con la Medicina Basada en la Evidencia presentó un nivel de fiabilidad bueno a excelente.

Los resultados de este estudio fueron publicados en la revista Biomedcentral Research Notes en el año 2012: *"Level of Evidence in Hand Surgery"* BMC Res Notes.2012; 5:665. <http://www.biomedcentral.com/1756-0500/5/665>.

El artículo original se adjunta en el anexo B.

Tabulación Cruzada de "Tipo de Revista" y "Nivel de Evidencia"						
NIVEL DE EVIDENCIA						
Revista	I	II	III	IV	V	Número Completo de Trabajos
RICMA (IC 95%)	0,0% (N.A) 0	0,0% (N.A) 0	16,7% (7,3; 26.1) 10	80,0% (69,9;90,1) 48	3,3% (N.A) 2	60
JHSE (IC 95%)	0,9% (N.A) 4	5,0% (3.02;6.9) 23	11,1%  51	82,4%  380	0,7% (N.A.) 3	461
JHSA (IC 95%)	8,3% (5,6;10.9) 34	10% (7.1;12.9) 41	14,1% (10.7;17,4) 58	67,6 % (63.1;72.1) 278	0% (N.A.) 0	411

RICMA: Revista Iberoamericana de Cirugía de la Mano/ JHSE: Journal of Hand Surgery European Volume/ JHSA: Journal of Hand Surgery American Volume/ IC: intervalo de confianza/ N.A. no aplicable: (cuando el número de proporción observado fue menor de 5/n).



## **3.2 Hipótesis Segunda: Estudio de las propiedades psicométricas de la versión española del DASH en la valoración de resultados de pacientes operados del Síndrome del Túnel Carpiano**

Los resultados que se obtuvieron fueron los siguientes:

- 1) Los 42 pacientes completaron el estudio. 36 mujeres y 6 hombres, con una edad media de 54 años (rango de 34-63).
- 2) La media y la SD de los resultados pre y post - quirúrgicos y la diferencia entre los resultados pre y post-quirúrgicos para cada medida están descritos en la tabla.
- 3) La sensibilidad al cambio para cada medida se muestra como número positivo si hay mejoría y como valor negativo si hay empeoramiento como resultado.

El cuestionario DASH presentó un ES de 0.68 y una SRM de 100. Las medidas del examen físico fueron menos sensibles, con resultados que abarcan desde 0.35 (SRM) y 0.30 (ES) para la destreza manual a 0.00 (SRM) (ES) para la fuerza en la pinza. La fuerza de puño y balance articular de la muñeca dio valores negativos (Tabla). La sensibilidad al cambio de los resultados del SF-36 fueron 0.07 (ES) y 0.08 (SRM) para el SF-36 bp y 0.14 (ES) y 0.21 (SRM) para el SF-36rp. El Cuestionario del Túnel Carpiano tuvo la mayor sensibilidad al cambio (rango 0.51- 1.4). (Tabla1, Fig 1).

Los resultados demuestran que la sensibilidad para detectar un cambio de la versión española del DASH es mayor que la exploración física y el instrumento genérico Sf-36rp y SF-36bp, pero menor que el instrumento específico CTS , para el tratamiento del STC.

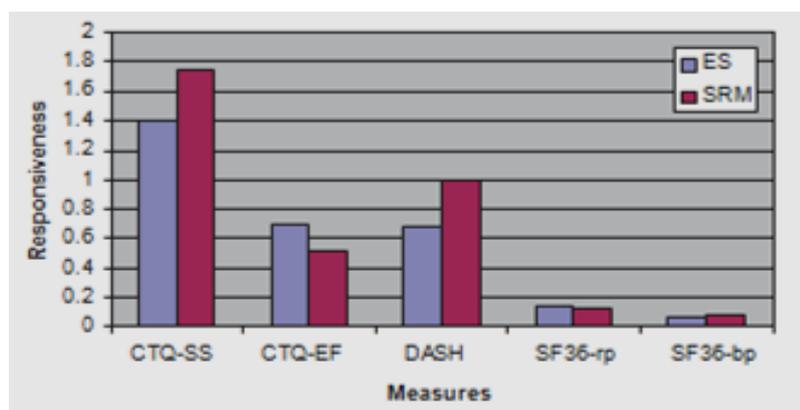
Tabla 1: Responsiveness del DASH, CTS, SF-36 y medidas del examen físico

Measure	Pre-op mean (SD)	Postop mean (SD)	Pre-post mean (SD)	E.S.	SRM.
CTQ-SS	3.26 (0.92)	1.96 (0.68)	1.30 (0.74)	1.41	1.75
CTQ-EF	2.94 (0.84)	2.35 (0.71)	0.59 (1.14)	0.7	0.51
DASH	51.01 (19.62)	37.63 (51.01)	13.37 (13.25)	0.68	1.00
SF-36rp	20.31 (33.19)	25 (44.72)	4.69 (38.96)	0.14	0.12
SF-36bp	42.44 (22.63)	44.06 (21.59)	1.63 (18.87)	0.07	0.08
Dexterity time	5.94 (3.56)	4.48 (1.44)	1.06 (3.06)	0.30	0.35
Grip (kg)	26 (7.46)	24.88 (7.41)	1.12 (4.51)	0.15	0.25
P-P pinch (kg)	5.71 (1.49)	5.97 (1.54)	0.26 (1.16)	0.17	0.22
Key pinch	6.35 (1.84)	6.35 (1.21)	0.00	0.00	0.00
ROM	203.53 (27.43)	195.53 (19.91)	8.00 (15.26)	0.29	0.52
2PPD (mm)	5.53 (1.18)	5.35 (1.00)	0.18 (1.29)	0.18	0.14

La puntuación del DASH va desde 0 (no discapacidad, buena función) a 100 (muy severa discapacidad, mala función); el SF36 va desde 0 (la peor salud) a 100 (la mejor salud); para el CTQ, puntuación alta significa peor resultado. Pre-op= medidas

antes de la cirugía; postop= medidas a los 3 meses de la cirugía, pre-post= cambios en la puntuaciones desde antes de la cirugía a los 3 meses después, SD: desviación estándar, E.S.= tamaño del efecto, S.R.M= medida estandarizada de la respuesta.

Fig 1: Validez de construcción del DASH. La sensibilidad al cambio del DASH (instrumento específico de la extremidad superior) es más alta que la del SF-36 (instrumento genérico) y más baja que el instrumento CTS (CTQ). ES= effect size (tamaño del efecto); SRM = standardised response mean (media estandarizada de la respuesta).



Los resultados de este estudio fueron publicados en el año 2009 en la revista *The Journal of Hand Surgery (European Volume)*: "*The Relative Responsiveness and Construct Validity of the Spanish Version of the DASH Instrument for Outcomes Assessment in Open Carpal Tunnel Release*". *JHSE* 2009; 34:72-5 . El artículo original se adjunta en el anexo C.

### **3.3 Hipótesis tercera: Estudio de las propiedades psicométricas de la versión corta española del Instrumento 6-CTS**

Los resultados que se obtuvieron con el análisis estadístico fueron los siguientes:

- 1) En este estudio no hubo ninguna pérdida de ítems en ninguno de los 3 instrumentos PRO utilizados.
- 2) Confiabilidad: el coeficiente alfa de Cronbach fue 0.81 y el ICC fue 0.85 ( $p < 0.001$ ). La diferencia media en la puntuación del CTS-6 para las dos administraciones fue 0.09 (95% de intervalo de confianza - 0.07 a 0.26,  $p = 0.269$ ) (Tabla 1 )
- 3) Validez de Construcción: hubo una fuerte correlación positiva entre la puntuación del CTS-6 y el QuickDASH ( $r = 0.685$ ,  $p < 0.001$ ), y una moderada negativa correlación con el EQ-5D ( $r = -0.45$ ,  $p = 0.003$ ).

Los resultados de este trabajo han demostrado que el instrumento PRO 6-CTS presentó un buen nivel de consistencia interna (coeficiente alfa de Cronbach de 0,81) y una alto nivel de fiabilidad (test-retest de 0,09) y con un excelente nivel de correlación intraclase (ICC= 0,85). Además, las correlaciones fueron concordantes con la hipótesis formulada, lo cual es una prueba de una buena validez de construcción.

Tabla 1: Resultados del análisis de la fiabilidad (N=40)

T-Retest Reproducibility		Internal Consistency	
Mean Score Difference (95% CI)*	Intraclass Correlation Coefficient (95% CI)*	Cronbach Alpha Coefficient	
0.09	0.85	0.81	
(- 0.07 - 0.26)	(0.71-0.92)		
P=0.269	P<0.001		

\* Diferencia media entre las puntuaciones y la correlación de dos medidas sucesivas separadas 1 semana.

CI: intervalo de confianza

- 4) Sensibilidad al cambio (responsiveness): Nuestro estudio, con 40 pacientes, mostró que el instrumento 6-CTS tenía alta sensibilidad al cambio con un ES de 3.06 y una SRM de 2.94, el QuickDASH presentó un nivel más bajo de sensibilidad al cambio con un ES de 1.62 y un SRM de 1.34. El EQ-5D presentó el nivel más bajo con un ES de 0.76 y un SRM de 0.79. (Tabla 2).

Tabla 2: Sensibilidad al cambio (responsiveness)

N=40	Pre-Op Media (SD)	Post- Op Media (SD)	Cambio pre-postop Media SD	ES	SRM
CTS-6	3.65 (0.74)	1.37 (0.43)	-2.27 (0.77)*	3.06	2.94
Quick DASH	56.34 (20.48)	23.01(13.18)	-33.33 (24.85)*	1.62	1.34
5Q-5D	0.56 (0.30)	0.80 (0.19)	0.23 (0.29)*	0.76	0.79

\* la puntuación media entre 1 sem antes de la cirugía y 3 meses después fue significativa ( $p < 0.001$ ). (SD) Desviación estándar. (n) tamaño de la muestra

En este estudio se demuestra la confiabilidad, validez de construcción y sensibilidad al cambio de la versión española del cuestionario específico para medir resultados en el síndrome del túnel carpiano, lo que convierte a este cuestionario en un instrumento fiable, cómodo y práctico tanto como para el paciente como cirujano para determinar resultados en una de las patologías que más gastos ocasionan al Sistema Sanitario, tanto por el uso de los servicios sanitarios como por las bajas laborales que ocasiona.

Los resultados de este estudio están en fase de revisión para publicación bajo el título *"Reliability and Construct Validity of the Spanish version of the 6-item CTS symptoms scale for outcomes assessment in Carpal Tunnel Syndrome"*.

El artículo original se adjunta en el Anexo D.

# 4

## DISCUSIÓN

### 4.1 Hipótesis Primera: Estudio del Nivel de Evidencia en Cirugía de la Mano

Los resultados de esta hipótesis han demostrado con un nivel de fiabilidad bueno-excelente que la variable "nivel de evidencia científica" es dependiente de la variable "tipo de revista". El uso de Kappa es importante, ya que nos permite conocer el grado de acuerdo intra-observador e inter-observador teniendo en cuenta el azar. Un coeficiente Kappa estadísticamente significativo indica que el acuerdo es diferente a cero (acuerdo nulo). Sin embargo, la interpretación de valores Kappa obtenidos es subjetivo, y han sido propuestas diferentes clasificaciones o guías de interpretación del

coeficiente Kappa para el análisis de la fiabilidad. En nuestro resultado, el nivel de acuerdo de análisis inter-observador y el intra-observador ha mostrado un valor Kappa comprendido entre 0.617 y 0.751, que puede ser considerado que tiene un nivel de fiabilidad de bueno a excelente (Landis 1977, Silman 1995), en la evaluación del nivel de evidencia y el tipo de revista.

Resultados similares han sido publicados antes. Obremskey y cols. (Obremskey 2005), en el estudio del nivel de evidencia de revistas de ortopedia, han publicado valores Kappa de 0.62 para acuerdos inter-observadores entre revisores de poca experiencia, y valores Kappa de 0.75 para inter-observadores experimentados. Ningún análisis intra-observador fue publicado por estos autores.

No hay muchos estudios sobre el nivel de evidencia científica en revistas de cirugía de la mano o en campos similares de investigación, como revistas de cirugía ortopédica y plástica. Sinno y cols., revisaron 726 artículos de 6 revistas diferentes de cirugía plástica y el nivel de evidencia fue evaluado usando una clasificación basada en el Centro para el Nivel de Evidencia de Oxford (CEBM). Hanzlik y cols (Hanzlik 2009) evaluaron 551 artículos del Volumen Americano del JBJS de los años 1975 (134 artículos), 1985 (123 artículos), 1995 (120 artículos) y 2005 (174 artículos). El nivel de evidencia fue evaluado siguiendo una clasificación incluida en la guía de autores (sistema de clasificación JBJS), la cual es muy similar a la desarrollada por la CEBM, con el fin de demostrar la tendencia del nivel de evidencia durante 30 años. Obremskey y cols (Obremskey 2005) revisaron 382 artículos de investigación clínica de nueve revistas diferentes con el fin de evaluar el nivel de evidencia científica en revistas de cirugía ortopédica. En la presente revisión fueron evaluados 932 artículos de tres

revistas específicas de cirugía de la mano, lo cual supone la población más amplia estudiada hasta ahora, para evaluar el nivel de evidencia científica.

Desde que fue descrito el primer sistema de clasificación del nivel de evidencia científica de los artículos de investigación clínica (Sackett 1986), la Medicina Basada en la Evidencia ha llegado a ser una parte importante de la práctica clínica. Los cirujanos de mano debemos entender el nivel de evidencia para llegar a ser conocedores de la veracidad y la utilidad de los datos que proporcionan los artículos consultados. Pocas investigaciones se han hecho para analizar el nivel de evidencia científica en revistas relacionadas con cirugía de mano en relación con otros campos de investigación clínica afines, como en revistas de cirugía ortopédica (Hanzlik 2009) y cirugía plástica (Sinno 2011). Sólo una revista de cirugía de mano ha sido analizada para ver su nivel de evidencia en un periodo de seis meses, y ha sido comparada con otras de publicaciones ortopédicas (Obremskey WT 2005). Hasta nuestro conocimiento, no existen trabajos previos publicados en lengua castellana que comparen el nivel de evidencia científica de la RICMA con otras revistas científicas del campo de cirugía de la mano.

Los resultados de este estudio demuestran que la mayoría de los artículos clínicos publicados en cirugía de la mano tienen muy poco nivel de evidencia científica (80% nivel IV en la JHSE o RICMA y 67.6% en la JHSA). La mayoría de los artículos fueron series de casos y menos frecuentemente estudios cohortes y casos-control de pobre calidad comparado con revistas científicas de cirugía ortopédica (48% estudios de nivel IV (Hanzlik 2009), plástica (40% estudios de nivel IV) (Sinno 2011) y de oftalmología (58% estudios de nivel IV) (Lai 2006). Sin embargo, otras revistas de cirugía, como las relativas a otorrinolaringología, presentan un porcentaje similar a la JHSE y RICMA (80% estudios de nivel IV) (Bentsianov 2002). El porcentaje de

artículos de nivel IV en la JHSA fue menor, comparado con el resto de revistas de cirugía de mano investigadas, y fue muy similar al del trabajo publicado por Obremsky y cols. (Obremsky 2005), donde se notifica la existencia de un 68.8% de artículos de nivel IV, en una revisión de 32 artículos publicados en la JHSA de Enero a Junio de 2003.

El porcentaje de artículos con un alto nivel de evidencia científica (nivel I y II) fue mayor en la JHSA (8.3% nivel I y 10% nivel II), comparado con la RICMA (0%) y la JHSE (0.9% nivel I y 5% nivel II). Mientras que si se compara con otras revistas, hubo un 21% de nivel I y un 15% de nivel II de evidencia en revistas de cirugía ortopédica, 3% de nivel I y 16% de nivel II en cirugía plástica (Sinno 2011), 18% de nivel I y 8% de nivel II en revistas de oftalmología (Lai 2006), y 7% de niveles I y II en otorrinolaringología (Bentsianov 2002).

El porcentaje de artículos con nivel III (mayoritariamente estudios de casos-control y cohortes no consecutivos o con poblaciones muy limitadas) publicados en la RICMA (16.7%) fue mayor comparado con la JHSE (11%) y la JHSA (14%); y similar a otras revistas: 16% en revistas de cirugía ortopédica (Hanzlik 2009), 16% en otorrinolaringología (Bentsianov 2002), y 16% en oftalmología (Lai 2006).

En el 2004, se publicó un estudio sobre la calidad de los ECA relacionados con el tratamiento de diferentes patologías en la extremidad superior, publicados en 4 revistas de cirugía de mano y de ortopedia, durante un periodo de 11 años. La calidad de los artículos fue evaluada siguiendo los parámetros de la escala de Jadad modificada que consiste en 3 ítems (randomización, enmascaramiento y pérdidas en el seguimiento). A puntuación más alta (0-5) mayor calidad. De los 92 artículos publicados como ensayos clínicos aleatorizados, 40 describían un correcto método de randomización, 31 artículos no describían el método, y 21 artículos describían una

inapropiada aleatorización. La puntuación media de calidad según la escala de Jadad fue de 2. Los resultados de este estudio sugieren que la calidad de los ECA publicados relacionados con la patología del miembro superior deben mejorar siguiendo la guía CONSORT para valorar su calidad.(Gummesson 2004)

Algunos autores han criticado el escaso número de artículos de alto nivel de evidencia en cirugía (Horton 1996). Aun así, las críticas puede que parezcan exageradas si se tiene en cuenta que un ensayo clínico aleatorizado quirúrgico es diferente de un ensayo que compara un medicamento con un placebo. Los procedimientos quirúrgicos son invasivos (herida quirúrgica); es difícil aleatorizar pacientes, el enmascaramiento es un problema en un ensayo quirúrgico, y ambos son muy caros pero necesarios. Por ejemplo, en un ECA reciente se valoró el resultado de la artroscopia de rodilla en la osteoartritis. En el grupo control al que no se realizó la artroscopia se le realizó las mismas incisiones de la artroscopia. El efecto del enmascaramiento fue evidente cuando el grupo control y el grupo artroscopia tuvieron el mismo resultado de mejoría. Si no se tienen estudios aleatorios de alta calidad, no se puede tener una revisión sistemática que sintetice la evidencia anteriormente descrita. Por otro lado cada cirujano tiene preferencia por una técnica quirúrgica en particular, por ejemplo, si un cirujano realiza de manera habitual una transposición simple subcutánea del nervio cubital para el tratamiento del síndrome del túnel cubital, sería poco ético para él participar en un ensayo clínico aleatorizado y realizar otras cirugías a los pacientes para el mismo problema.

Debemos comenzar a pensar y usar la Medicina Basada en la Evidencia. Establecer los resultados de nuestras intervenciones en el contexto de una nueva revisión sistemática. Nosotros podemos aprender de los terapeutas de mano. Sus revistas están llenas de revisiones sistemáticas de patologías comunes como la eficacia

del vendaje en la epicondilitis lateral, la efectividad de los ejercicios de manos en pacientes con artritis reumatoide, la rehabilitación de pacientes con epicondilitis lateral, manejo de la patología del manguito rotador etc. El fallo de no encontrar una buena evidencia, debe ser el trampolín que nos impulse a realizar una investigación apropiada. Debemos hacer un continuo esfuerzo para mejorar nuestro conocimiento médico usando la metodología científica y participar en revisiones sistemáticas con nuestros resultados en la práctica médica diaria.(R.M. Szabo 2008). Una revisión sistemática no es aquella en la que un experto presenta su punto de vista y lo apoya en la bibliografía. Por el contrario, se trata de una recogida de bibliografía según un esquema organizado para después valorarla en comparación con un determinado estándar. Es decir, una RS no necesariamente debe ser escrita por un experto, y de hecho no debería ser así. De las RS se extraen directrices que satisfacen los pasos más importantes de la MBE: la formulación de una pregunta que pueda responderse, la recogida de evidencias y su valoración. Nos ofrecen una revisión completa, sin obligar a consultar personalmente la bibliografía primaria.

Sin embargo, hoy en día existe mucha resistencia a la práctica de la Medicina Basada en la Evidencia. Suele verse entre médicos que conocen su teoría pero no la practican, y pueden estar molestos por creer que la práctica de la MBE exige un alto nivel de conocimientos matemáticos. En realidad, la MBE no precisa de un especial manejo de los números. El médico que practique la MBE no necesita ser estadístico; sólo debe saber cómo utilizar lo que los estadísticos producen. Mayor resistencia pueden presentar los médicos que piensan que la MBE es un pretexto que limita su autonomía profesional, es decir, los que piensan que bajo las normas de la MBE, los médicos quedan reducidos a ser meros autómatas y que la MBE no es práctica. Dichas resistencias son legítimas, puesto que los médicos tienen cada vez menos tiempo libre,

por lo que la lectura y valoración de la bibliografía les hace consumir mucho tiempo. (Bernstein 2004)

La mayoría de artículos publicados en revistas de cirugía de mano, como hemos visto, son series de casos que incluyen 20 o incluso menos pacientes. Está claro que un sólo caso o una pequeña serie no puede ser representativa de una población mayor. Sin embargo, interpretado en un contexto apropiado, se puede obtener información que puede ser útil. El error más frecuente es que las conclusiones no están basadas en datos validados. Los editores son los que tienen el papel más importante para asegurarnos que la metodología usada en sus cortos seguimientos es correcta, que ofrecen alternativas en sus interpretaciones de los datos y que informan al lector de nuevas hipótesis que se pueden plantear para futuras investigaciones. Por otro lado, los diseños de caso-control suelen ser usados para investigar posibles causas. Si un factor de riesgo tiene una fuerte asociación con una enfermedad, podría establecerse la hipótesis de que dicho factor causa la enfermedad (odds ratio  $> 1$ ), y lo contrario, cuando existe una leve asociación (odds ratio  $< 1$ ). Existen una gran variedad de condiciones en cirugía de la mano que pueden ser abordadas para su estudio usando el diseño de caso - control. Algunos ejemplos serían, la asociación entre diferentes tipos de trabajo y el desarrollo del síndrome del túnel carpiano, diferentes factores que afecten tanto positiva como negativamente a cirugías reconstructivas como las artrodesis de muñeca, o la necesidad del uso de antibióticos profilácticos en la mano traumática. Asimismo, también frecuentes en las revistas de cirugía de mano, son los estudios de corte de sección transversa. En este caso, la muestra es estudiada en un punto de tiempo determinado, no hay seguimiento. Se puede estudiar la prevalencia del factor estudio, pero al no haber seguimiento no podemos establecer su incidencia. La principal ventaja de estos diseños es que son muy baratos, pero sólo pueden

estudiarse aquellas variables que sean relativamente frecuentes, ya que condiciones poco frecuentes requerirían muestras de mayor tamaño siendo difícil establecer una relación con el factor a estudio. Este diseño, como los anteriores, nos sirven para realizar estudios de mayor evidencia científica. (B. Graham 2005)

Y, ¿qué hacer cuándo la MBE no es concluyente? El cirujano recurre principalmente a su experiencia ("en mis manos", "familiaridad con el tratamiento", "lo que me han enseñado") más que a la propia perspectiva del paciente, exceptuando, las opciones de menos complicaciones y más rápida recuperación, lo cual beneficia tanto al paciente como al médico. Sin embargo, no nos debemos olvidar de la expectativa del paciente. Para un paciente que sufre un determinado proceso con la consiguiente modificación de su calidad de vida, la eficacia del tratamiento que recibe le aporta cierta satisfacción. Esta satisfacción representa el grado en que la atención prestada, durante el proceso, satisface las expectativas del paciente para mejorar y/o recuperar su calidad de vida. La más impecable y arriesgada de las intervenciones quirúrgicas puede no ser apreciada por un paciente que no ha establecido una buena relación con su cirujano. (Donebian 1989). En ocasiones la toma de decisiones resulta difícil porque los médicos se enfrentan al dilema de basarlas en informaciones imperfectas o inadecuadas. La existencia de verdades no comprobadas puede atribuirse a la falta de recursos; es decir, no hay tiempo ni dinero suficiente para comprobarlo todo. El problema es que incluso con recursos ilimitados todavía habrá algunas afirmaciones ciertas cuya verdad no pueda establecerse. Esto precisamente fue demostrado en 1930 por Kurt Gödel en su sorprendente teorema de la "incompletitud". Una traducción en lenguaje común del teorema dice que, "no todas las afirmaciones verdaderas son probablemente verdaderas". Por lo tanto nos quedan dos opciones poco atractivas: la imperfección por un lado y la inadecuación por el otro. En la práctica no

podemos confiar en encontrar para cada caso la información perfecta. En realidad, las decisiones hay que tomarlas con la información proveniente de nuestra experiencia y de la búsqueda externa. Siempre hay que tomar una decisión y para realizar una elección con sabiduría y no dejarla a la suerte, Sackett y Haynes aconsejan 5 pasos fundamentales para la utilización de la Medicina Basada en la Evidencia: 1) formular preguntas que puedan responderse, 2) reunir evidencias, 3) valorarlas, 4) ponerlas en práctica y 5) evaluar el proceso.

Ningún análisis de tendencia es una limitación de este estudio, y dicha información debería ser el propósito de futuros estudios, en el sentido de entender cómo la evidencia científica publicada en revistas de cirugía de la mano ha cambiado, y cómo la relación entre cambios en el nivel de evidencia y cambios en el índice de impacto ha variado también con el tiempo. Queda mucho trabajo por hacer. Por ejemplo, a pesar del incremento en el uso de la artroscopia de muñeca, la eficacia de las intervenciones sólo ha sido estudiada en 4 estudios randomizados con una puntuación media en la escala de Jadad de 0,5 (rango de 0-1). Esto contrasta con los 50 estudios randomizados que valoran los resultados de intervenciones realizadas usando el artroscopio de hombro con una puntuación media en la escala de Jadad de 3.0 (Tadjerbashi 2014)

Después de revisar una enorme cantidad de artículos publicados en revistas de diferentes partes del mundo, otras cuestiones afloran, tales como si las diferencias encontradas son un reflejo de diferentes prioridades regionales, o de qué manera las inversiones utilizadas para la investigación tienen un impacto en los resultados, e incluso si diferentes países son los principales contribuidores en estudios de alto nivel.

The Journal of Hand Surgery ( volumen americano) añadió en el año 2008 una sección de Revisión y Medicina Basada en la Evidencia. Esta sección revisa los

artículos publicados con el fin de promover el papel de la investigación en cirugía de la mano (W.Hammert 2013), es quizás una de las causas del aumento del número de artículos publicados con mayor nivel de evidencia en comparación con la Revista RICMA y Revista JHSE, como hemos demostrado en nuestra hipótesis (el nivel de evidencia de los artículos publicados es dependiente de la revista en la que se publica).

---

## **4.2 Hipótesis Segunda: Estudio de las propiedades psicométricas de la versión española del DASH en la valoración de resultados de pacientes operados del Síndrome del Túnel Carpiano**

Los resultados de este trabajo demuestran que la sensibilidad para detectar un cambio de la versión española del DASH es mayor que la exploración física y el instrumento genérico Sf-36rp y SF-36bp, pero menor que el instrumento específico CTQ , en la valoración de resultados tras tratamiento del STC. Podemos discutir este concepto en los siguientes apartados:

### 1) Equivalencia intercultural y sensibilidad al cambio del DASH:

El objetivo último en la valoración de resultados es utilizar medidas estándares y equivalentes conceptualmente a través de diferentes culturas y lenguas. Así, la idea de poder comparar resultados con estudios previamente realizados o en desarrollo, es una realidad hoy en día gracias a la existencia de instrumentos de medición ampliamente difundidos como el DASH (Disability of the Arm, Shoulder and Hand). El DASH fue desarrollado por la AAOS (American Academy of Orthopaedic Surgeons) y el IWH (Institute for Work and Health, Canadá) para medir discapacidad y síntomas en relación con los MMSS y sus afecciones músculo-esqueléticas. El instrumento DASH había sido desarrollado en un país con diferente lengua y cultura. De esta manera, el

cuestionario debía ser adaptado al castellano antes de su uso con el fin de obtener una versión en castellano equivalente conceptualmente a la original americana. Una vez adaptado el cuestionario, se debían analizar sus propiedades psicométricas; confiabilidad (reliability), sensibilidad al cambio (responsiveness) y validez de construcción.

La confiabilidad de la versión española del DASH fue analizada midiendo la consistencia interna de la escala del instrumento STC y el DASH con un resultado de coeficiente alfa de Cronbach  $>0.7$ , lo que indicaba una buena confiabilidad. Esto indicaba que todos los ítems entre cada escala del instrumento CTS y entre los ítems del DASH estaban bien correlacionados. (Rosales 2002). Por lo tanto, ya sólo nos quedaba para demostrar la completa equivalencia de la versión española del DASH, la validez de construcción y la sensibilidad al cambio.

Diferentes versiones estándares del DASH no han publicado el análisis de la sensibilidad al cambio en las publicaciones originales (Lee 2005, Orfale 2005, Padua 2003, Rosales 2002, Themistocleus 2006). La versión española del instrumento DASH presentó un buen nivel de sensibilidad al cambio a las 12 semanas después de la liberación del túnel carpiano con un ES de 0,68 y un SRM de 1.00. Un nivel similar de sensibilidad al cambio ha sido publicado por otros autores: Kotsis y Chung (2005) han publicado un SRM con una media de 0.7 (versión Americana); De Smet et al (2007) (versión Sueca), presentó un ES de 0.87 y un SRM de 0.69; Greenslade et al (2004) (versión Británica) mostró una SRM de 0.62; y Gay et al.(2003) (versión Americana) mostró una ES de 1.01, y una SRM de 1.13. Gummenson et al(2003) con una versión Sueca demostró a los 6 meses tras la cirugía una ES de 0.7, y una SRM de 1.0. Una tasa más baja de sensibilidad al cambio del DASH (ES de 0.49, SRM de 0.43) ha sido

publicado en la población Británica con STC por Hobby et al.(2005); sin embargo, una alta tasa de abandono (25%) en el estudio hace que esos resultados sean inadecuados para la comparación intercultural. La versión española del DASH detecta un mediano-gran cambio de significancia clínica después de la cirugía del túnel carpiano similar a otras versiones estándares y equivalentes.

Uno de los estudios a destacar y el primero en analizar si el DASH tenía suficiente capacidad de detectar cambios clínicos tras cirugía del túnel carpiano fue el realizado por Gay y Amadio (Gay 2003). La sensibilidad al cambio del DASH se demostró que era comparable al instrumento CTS en la valoración de patología en muñeca y mano en general, pero no había sido aún validado su sensibilidad al cambio en relación al STC. Para responder a esta pregunta, se comparó la sensibilidad al cambio tras la cirugía, del DASH, CTS y SF36, comparando su ES y SRM. Como resultado se obtuvo que el instrumento CTS era el más sensible y el SF-36 el menos, quedando el DASH en posición media. Aunque el DASH es menos sensible que la escala de severidad de síntomas del CTS a las 6 y 12 semanas, es igual que la escala funcional del CTS. A pesar de esto, tanto el DASH como la escala funcional del CTS presentan sólo una sensibilidad moderada a las 6 semanas, tiempo en el que la mayoría de los pacientes se incorporan a su trabajo y sus actividades de la vida diaria.

La sensibilidad al cambio es sólo uno de los diferentes factores que son considerados cuando elegimos un instrumento de valoración de resultados. Algunos investigadores definen "responsiveness" no sólo como la capacidad de detectar el cambio (sensibilidad) sino como la capacidad de detectar un cambio significativo clínicamente (relevancia). Grandes cambios pueden no ser significativos y en cambio pequeños cambios pueden ser muy significativos. En este estudio, se valoró la satisfacción del paciente como un cambio clínicamente importante a estudiar. Para

muchos autores, la satisfacción del paciente no es un dato que tenga que ser incluido en los instrumentos de valoración, ya que existe mucha variación de la importancia que el paciente le da y no se puede controlar. Aún así, la información recogida con el instrumento CTS, refleja el estado de satisfacción del paciente más que otras medidas físicas. La tasa de satisfacción en este estudio mostró un coeficiente de Cronbach de 0.95, indicando una consistencia interna excelente. En cambio, poca correlación se observó entre la puntuación de satisfacción, el DASH y el cambio de resultados en el CTS, que se debió a la pequeña variación del grado de satisfacción de los pacientes a las 6 semanas (85% completamente satisfechos) respecto a las 12 semanas.

Se demostró que la sensibilidad al cambio del DASH tiene la suficiente magnitud como para que pueda ser usado como instrumento de valoración de resultados tras cirugía del STC cuando ésta es superior a las 12 semanas. Aunque los estudios que utilicen el DASH necesitan mayor tamaño de muestra que aquellos en los que se usa el instrumento CTS, el DASH permitirá una comparación directa del impacto del STC con otras patologías del miembro superior.(Gay 2003)

## 2) Efecto del tratamiento y resultados:

El análisis de la sensibilidad en detectar un cambio significativo, permite conocer qué medida debería ser usada para valorar resultados de cualquier tratamiento en una población específica, y calcular el tamaño de la muestra para cualquier ensayo clínico (Gay 2003, Liang 1995). Los resultados han demostrado que la magnitud del efecto del tratamiento detectada por el DASH era mayor que la obtenida por las medidas del examen físico en la valoración de resultados en la cirugía del STC. Este concepto tiene importantes implicaciones en la interpretación de cualquier estudio de seguimiento en la cirugía del túnel carpiano. El uso de medidas con baja sensibilidad al

cambio puede llevar a falsas conclusiones acerca de la efectividad de un tratamiento. En este trabajo, medidas tales como el rango de movilidad y la fuerza de puño mostraron un valor negativo tras el tratamiento, mientras otras medidas como el DASH y el CTS, mostraron mejoría tras el tratamiento.

En los casos en que el propósito del trabajo sea el análisis de resultados siempre vamos a tener medidas antes y después del tratamiento, siendo la exposición al tratamiento la variable independiente. Así, podemos calcular la mejoría tras tratamiento, calculándose la magnitud de dicha mejoría clínica en base al efecto del tratamiento que se mide mediante el tamaño del efecto (ES= effect size) (media de la diferencia pre-post tratamiento / desviación estándar de la línea base) o la media estandarizada de la respuesta (SRM= standarized response mean) (media del cambio/desviación estándar del cambio). La variable independiente que define la entrada de pacientes al estudio es el tratamiento, mientras la variable dependiente es el resultado o mejoría, de ahí que esta variable también se denomine variable resultado o "variable outcome". El conocimiento de cuáles son las medidas más sensibles para valorar resultados en una población estudio específica es un paso fundamental en el campo de la investigación clínica. En caso contrario como se comentó anteriormente, podríamos llegar a conclusiones erróneas en nuestro estudio y observar malos resultados tras tratamiento debido a que a lo mejor estamos usando medidas que no son sensibles para detectar cambios de importancia clínica, cuando en realidad el tratamiento ha sido eficaz.

### 3) Validez de Construcción y análisis de la sensibilidad al cambio:

La validez de construcción es una expresión del grado en el que un instrumento mide aquello que se propone medir. Métodos comunes para obtener la validez de construcción incluyen un examen de las lógicas relaciones que deberían existir entre

las medidas obtenidas en un estudio con otras y/o entre patrones de puntuación entre grupos de individuos (Medical Outcomes Trust, 1997). El concepto de detectar cambios en el tiempo es parte de la valoración de la validez y por lo tanto puede ser definida como validez de construcción longitudinal (Gummenson 2003). De hecho, Gay et al.(2003), como hemos descrito previamente, demostraron en un estudio de cohorte longitudinal que el DASH presentó un nivel de responsiveness más alto que el SF-36bp y el SF-36 pf y un nivel más bajo que el instrumento específico CTS. La misma hipótesis fue establecida en este trabajo para demostrar la validez de construcción de la versión española del DASH. Los resultados coincidieron con esta hipótesis como prueba de la validez de construcción de la versión española del DASH. Los resultados de este trabajo presentan una buena evidencia de sensibilidad al cambio de la versión española del DASH para valoración de resultados de la liberación del túnel carpiano, similar a otras versiones equivalentes, mayor incluso que la exploración física.

El instrumento específico para valoración de resultados en el STC se ha demostrado que tiene la mayor capacidad de detectar cambios clínicos tras la cirugía del STC que el DASH y el Michigan Hand Outcomes Questionnaire (MHQ). La sensibilidad del CTS ha sido comparada con la del DASH en diversos estudios (Gay, Amadio 2003), (Rosales 2008) así como la del MHQ y el CTS. Es esencial que todos los instrumentos sean comparados respecto a su sensibilidad para de este modo proporcionar una plataforma basada en la evidencia en la cual el investigador pueda elegir el instrumento óptimo para su estudio. Además con el uso del instrumento más sensible se reduciría el número de participantes necesarios en el estudio. En un estudio con nivel de evidencia diagnóstica 1 publicado en el año 2009 (JS. Chatterjee 2009), se comparó la sensibilidad al cambio tras descompresión del túnel carpiano. Como resultado se demostró que el CTS es más sensible que el MHQ, por lo tanto si el clínico

quiere comparar estudios de valoración de resultados tras STC, se recomienda el uso del CTS, en cambio, si va a comparar pacientes con STC con otras patologías del miembro superior, el instrumento MHQ es el más favorable dado su carácter multidimensional atribuible a sus 6 dominios.



## **4.3 Hipótesis tercera: Estudio de las propiedades psicométricas de la versión corta española del instrumento 6-CTS**

El uso de instrumentos de salud específicos para medir la valoración de resultados está en crecimiento. La cirugía del síndrome del túnel carpiano es una de las más frecuentes realizadas hoy en día en un servicio de cirugía de mano, de ahí la importancia de valorar sus resultados.

La medición de resultados es la piedra angular de todos los trabajos de investigación. Un instrumento de medida debe reunir unos criterios claves para ser útil, siendo los más importantes una satisfactoria confiabilidad, validez y sensibilidad en detectar cambios clínicos. El instrumento además debería ser sensible (Feinstein 1987), entendido como "que tenga sentido común".

Una confiabilidad excelente es una característica absolutamente necesaria de cualquier instrumento de medición. Una pobre confiabilidad reduce el uso de cualquier escala a pesar del grado de validez, es decir, si el instrumento no puede medir lo que se propone de una manera reproducible, tiene poco valor. No obstante, la confiabilidad por sí misma, aunque es una condición necesaria, no es suficiente para el uso de un instrumento si no tiene además una buena validez. (Badalamente 2013)

Cuando medimos resultados con instrumentos basados en la opinión del paciente, los instrumentos específicos han sido diseñados para mostrar mínimas diferencias pero clínicamente importantes en condiciones específicas. Sin embargo, estos instrumentos no son test diagnósticos. El diagnóstico de STC debe ser hecho de manera independiente a una escala de puntuación. Una historia clínica y exploración

física son esenciales. El diagrama de Katz- Stirrat es un método sensible y específico que nos ayuda al diagnóstico del STC. Se ha demostrado que la probabilidad de un paciente en tener STC está relacionado con la distribución de síntomas que el propio paciente marca en el diagrama. Puede ser incluido en la primera consulta del paciente, ya que es un método cómodo, barato y repetible.

Normalmente los estudios de liberación del túnel carpiano han incluido el alivio de síntomas y la mejoría en la función para la valoración de resultados, pero no han usado medidas estandarizadas que demuestren su validez o confiabilidad. Los pacientes buscan sobretodo la mejoría en la severidad de sus síntomas y el volver a sus actividades en la vida diaria (estado funcional), por lo que estos puntos deben ser valorados en los resultados de nuestro instrumento. ( Levine 1993)

Una de las propiedades de medición de los instrumentos PRO incluidos en el Medical Outcomes Trust en los criterios de revisión de los instrumentos ha sido "la carga para los encuestados", que lo define como el tiempo, energía, y otras demandas que ocurren en aquellas personas a quien se les administra el instrumento PRO. Para disminuir la carga de los encuestados la tendencia actual es reducir el tamaño de los instrumentos. Levine en 1993 publicó el instrumento específico CTS basado en una escala de 11 ítems para valoración de la severidad de síntomas y una escala de 8 ítems para valorar el estado funcional. En el año 2011 (Atroshi 2011) se publicó la escala de valoración de severidad de síntomas con 6 ítems que demostró una buena consistencia interna, confiabilidad y validez, comparada con la escala de 11 ítems. Esta escala consiste en 6 ítems que preguntan sobre la frecuencia que las parestesias, hormigueos y el dolor aparecen durante el día y despiertan por la noche. Así mismo, para completar la valoración de resultados tras cirugía del túnel carpiano, se desarrolló la escala de 2 ítems de valoración del dolor en la palma de la mano. Estos dos cuestionarios, CTS-6 y

escala del dolor en la palma, se han convertido en una de las herramientas más prácticas en valoración de resultados y por lo tanto deberían estar disponibles en todos los idiomas. Considerando que la versión española de la escala de severidad de síntomas del instrumento CTS-11 ítems está disponible (Rosales 2002), la adaptación al español de la versión española no implica grandes cambios. Asimismo, para la adaptación española de la escala de dolor en la palma se utilizó la versión española de la escala de dolor corporal del SF-36.

Los resultados de este trabajo han demostrado que el instrumento PRO 6-CTS presentó un buen nivel de consistencia interna (coeficiente alfa de Cronbach de 0.81) y una alto nivel de fiabilidad (test-retest de 0.09) y con un excelente nivel de correlación intraclass (ICC= 0.85). Además, las correlaciones fueron concordantes con la hipótesis formulada, lo cual es una prueba de una buena validez de construcción.

En este estudio, la versión española del 6-CTS presentó un coeficiente alfa de Cronbach de 0.81 y una ICC de 0.85. Resultados similares han sido publicados con el 6-CTS original (Cronbach= 0.86, ICC= 0.95) (Atroshi 2009). La diferencia media entre las puntuaciones de dos administraciones en el 6-CTS original fue de 0.03 ( 95% CI-0.07 a 0.12) (Atroshi 2009). En nuestra muestra la diferencia media fue de 0.09 (95% CI -0.07 a 0.26). Comparando con la versión española más larga (11 ítems escala severidad de síntomas) , con el mismo intervalo de tiempo de una semana, los resultados también fueron similares (diferencia media = 0.18, CI=95% -0.16 a 0.53, p= 0.248) (Rosales 2002). Por lo tanto, la versión española del 6-CTS presenta un alto nivel de confiabilidad, similar a la del original 6-CTS y a la versión española 11-CTS.

Existen diferentes formas de analizar la validez de un instrumento de salud: validez de contenido, validez de criterio y validez de construcción (Medical 1997, Terwee 2012). La validez de un instrumento se define como la

capacidad que presenta un instrumento para medir el concepto que propone medir. De este modo, para analizar la validez de construcción, lo que hicimos fue establecer una relación lógica entre las medidas obtenidas con un instrumento con otras medidas de probada validez, mediante un análisis de correlación. Así, se incluiría la hipótesis que puede demostrar la teoría propuesta.(Mokkink 2010, Terwee 2012). Muchos factores tienen que ser considerados cuando elegimos una hipótesis. El factor más importante es la dimensión específica de salud, concepto que se entiende que debe ser medido usando un cuestionario completado por el paciente y la manera de valorarlo siguiendo un sistema de puntuación según el instrumento empleado. Atroshi demostró una validez convergente por una fuerte correlación entre la puntuación del original 6-CTS y la puntuación del QuickDASH ( $r=0.7$ ). En este estudio hemos observado un coeficiente de correlación muy similar ( $r=0.69$ ) entre la versión española 6-CTS y el QuickDASH. Además, se esperaba una correlación con nivel bajo-moderado negativo entre el 6-CTS y el EQ-5D ( $-0.46$ ) lo que constituye otra prueba de validez del 6-CTS (síntomas severos del STC se correlacionarían con niveles más bajos de calidad de vida ). Hasta nuestro conocimiento, ningún estudio previo ha usado el EQ-5D en el análisis de la valoración de construcción. Los resultados de este estudio demostraron que la hipótesis pre- especificada fue confirmada.

Basado en la escala de puntuación del COSMIN (Consensus- based Standards for the selection of health status Measurement Instruments) ( Mokkin 2010, Terwee 2012) para evaluar la calidad de estudios en las mediciones de las propiedades de los instrumentos PRO, el análisis de la validez de construcción de este estudio, reunió 10 de los 10 requerimientos estándares designados para una buena calidad metodológica: descripción de cómo fueron manejados los ítems perdidos, porcentaje de valores perdidos, cálculo del tamaño de la muestra, la hipótesis con respecto a las correlaciones

fue formulada a priori antes de la recogida de datos, los resultados esperados de las correlaciones fueron incluidos en la hipótesis, la magnitud absoluta o relativa de las correlaciones esperadas fueron incluidas en la hipótesis, hubo una descripción adecuada del instrumento comparador y sus propiedades, y finalmente se probó un adecuado diseño clínico y metodológico para la valoración de la hipótesis.

Nuestros resultados mostraron que el 6-CTS fue más sensible para detectar cambios clínicos que el instrumento Quick DASH y el Euroqol 5D.

Una de las ventajas de utilizar instrumentos de medida estandarizados es poder comparar los resultados. Hay 2 formas para comparar la sensibilidad al cambio de los resultados: la estadística y la heurística. La estadística incluye la interpretación de resultados de la sensibilidad al cambio de un instrumento mediante la medida del tamaño del efecto (ES) y la media estandarizada de la respuesta (SRM) que puede ser un hecho desafiante para los investigadores. Ambos parámetros se basan en determinar la cantidad de cambio en la desviación estándar para estimar el tamaño del efecto. Cohen (1998) definió el ES de 0.2 como pequeño, 0.5 como medio, y más de 0.8 como grande. Otros investigadores proporcionan la SRM, que a veces puede no ser clínicamente importante y difícil de interpretar. Por ejemplo, al valorar la fuerza de puño, una diferencia de 0.1 KG puede ser estadísticamente significativo pero no clínicamente importante. Por todos estos datos, a veces confusos, para el investigador de estudios clínicos en cirugía de la mano, es importante que sea capaz de interpretar los datos de los resultados. ( K Chung 2006). El otro camino para comparar la sensibilidad al cambio es la heurística, que se trata de preguntar al paciente y/o al proveedor si ha ocurrido un cambio; su magnitud, y la satisfacción del paciente con su estado. Este método heurístico de medir los resultados de la sensibilidad al cambio es importante para calcular el tamaño de la muestra y también para el investigador que lee

e interpreta los resultados de un estudio. A diferencia con los resultados estadísticos hay pocos estudios que lo usen para medir la sensibilidad al cambio en los resultados tras cirugía en el STC.

La diferencia mínima clínicamente importante (MCID) es un concepto perteneciente al método heurístico. El concepto de MCID es muy importante cuando consideramos el uso de instrumentos PRO para valoración de resultados primarios en un paciente. Beaton et al. resaltaron las diversas definiciones que los investigadores realizaban sobre la MCID, pero todas en esencia la resumían como la puntuación mínima necesaria de una medida que es clínicamente importante que debe ser obtenida del paciente tanto de manera positiva (resultado beneficioso para su salud) como negativa (resultado perjudicial para su salud), es decir el cambio más pequeño en el resultado de un tratamiento que un paciente identifica como importante. (Beaton 2002)

A través de los años, grandes pasos se han hecho para informar de lo que realmente importa en investigación clínica. En 1950, R.A. Fisher introdujo el concepto de significación clínica con el uso del valor "p" como punto de corte de 0.05; esto dio lugar como resultado, tanto a estudios con significación clínica como no significativos. Este valor de "p" tan rígido podía tener dos consecuencias: 1) diferencias clínicamente importantes podían ser observadas en un estudio como estadísticamente no significativas y por lo tanto ser ignorada como resultado de que la muestra sea pequeña (error tipo II) y 2) Incluso la diferencia más pequeña puede ser demostrada estadísticamente significativa al aumentar el tamaño de la muestra. Por lo tanto, la significación estadística no implica necesariamente importancia clínica. Con el fin de estudiar la importancia clínica, se introdujo el concepto de diferencia clínicamente importante (MCID) propuesto por Jaeschke et al. en 1989. MCID es el cambio más pequeño en un resultado que un paciente identificaría como importante. Por lo tanto,

MCID ofrece un umbral por encima del cual los resultados se experimentan como relevantes por el paciente que evitaría el problema de la mera significación estadística. Beaton et al, observaron que los investigadores habían definido el término de MCID de diferentes maneras, pero todas tenían el denominador común de significar como la mínima diferencia observable como clínicamente importante. (Beaton 2002). La identificación de la MCID en la puntuación de un instrumento es de gran ayuda para entender si el resultado de un instrumento representa realmente una mejoría del estado de salud del paciente. Para establecer esta mejoría debemos tener una referencia externa con la que compararla, que puede ser el grado de satisfacción del paciente, y/o la valoración del especialista.

Varias técnicas se han descrito para calcular la MCID. La más utilizada en nuestro campo es la llamada "Pregunta Ancla". Este método compara cambios en las puntuaciones con un "ancla" de referencia, que va a establecer si el paciente está mejor después del tratamiento en comparación al estado previo de acuerdo con la propia experiencia del paciente. En un punto específico en el tiempo después del tratamiento se le puede pedir al paciente que responda a la siguiente pregunta ¿Se siente que está mejor tras el tratamiento?. Como respuestas posibles usamos como herramienta el instrumento SF-36: mucho mejor, un poco mejor, casi igual, algo peor y mucho peor. Algo interesante para establecer la pregunta "ancla" sería el establecimiento de la misma antes de comenzar el tratamiento, preguntando al paciente qué resultado mínimo sería necesario para continuar con el tratamiento, por ejemplo, puede ser más importante el alivio de su dolor que mejorar su función. Sin embargo, el método "ancla" no es adecuado para las condiciones donde la mayoría de los pacientes mejoran y pocos permanecen sin cambios. Una posible solución a este problema es una variación en el cálculo de una puntuación de "beneficio clínico sustancial". Este cálculo no se basa en

los pacientes que mejoran versus los que no lo hacen, sino en los pacientes que mejoran y los que mejoran mucho.

La MCID varía de acuerdo a las enfermedades y los instrumentos de resultados, pero no depende de métodos de tratamiento. Por lo tanto, dos tratamientos diferentes para una enfermedad similar se pueden comparar con la misma MCID si el instrumento de medición del resultado es el mismo. También la MCID parece diferir en el tiempo después del tratamiento de la misma enfermedad. (Calfee 2012)

En el año 2006 se publicó un estudio para determinar la MCID en la escala de severidad de síntomas del cuestionario STC (Özyürekoglu Tuna 2006). Para ello pacientes diagnosticados de STC fueron tratados mediante infiltración con esteroides. Fueron valorados mediante el diagrama de Katz- Stirrat y completado el instrumento CTS una semana antes de la infiltración y 3 semanas después. Después del análisis estadístico, el resultado de la MCID en la puntuación de la escala de severidad de síntomas fue de 1.04. Este valor se considera importante considerando que el rango de puntuación va de 0 a 5. Para la práctica, se asume que una mejoría en la puntuación del resultado de 1.0 o más, representa una importante mejoría del estado del paciente. Este número será el valor de referencia para los investigadores cuando lean artículos sobre resultados con el instrumento específico CTS para valoración de la mejoría clínica del paciente, siendo la MICD, importante además para determinar el tamaño de la muestra. Se requieren más estudios para determinar si este valor de MCID puede ser usado en otros estudios clínicos que incluyan otras modalidades como la cirugía.

Ha sido demostrado que el cuestionario CTS es sensible para detectar cambios clínicos tras cirugía del STC. Un cambio clínico debe ser entendido como un resultado que cambia la perspectiva del paciente o del clínico. Por ejemplo, para el paciente puede ser volver al trabajo o dormir una noche entera, y para el clínico el retraso en la

indicación de cirugía. Estos estudios usan sobretodo el método estadístico para su evaluación. Levine, usó tanto el estadístico como el heurístico, midiendo la sensibilidad al cambio con el ES y SRM y comparando el cambio de la satisfacción del paciente tras la cirugía. La satisfacción del paciente se correlacionaba de manera muy alta con la mejoría en la escala de valoración de síntomas y moderadamente con la escala del estado funcional. Amadio et al, demostraron que el cuestionario CTS es más sensible a los cambios clínicos tras cirugía que el examen físico, hallazgos en el ENMG, instrumentos genéricos tales como el DASH o el SF 36.

El instrumento CTS fue introducido por Levine que publicó una SRM de 1.1 para la escala de severidad de síntomas y de 0.7 para la escala de estado funcional en 26 pacientes a los 3 meses de la cirugía. Amadio publicó en 22 pacientes una SRM de 1.8 para la escala de síntomas y de 1.3 para el estado funcional. Katz publicó un estudio randomizado con pacientes intervenidos por endoscopia y liberación abierta, con una SRM de 2.0 para la escala de severidad de síntomas y de 1.0 en la escala funcional. (Atroshi 1999). Como instrumento de valoración general en el estudio publicado por Atroshi se usó el SF-36, en nuestro estudio el EQ-5D. Ambos instrumentos miden la calidad de vida del paciente en este caso relacionado con el síndrome del túnel carpiano. Comparando el SF-36 con la población general, se evidenció que los pacientes con STC tenían una peor calidad de vida. Estos resultados son muy importantes a la hora de evidenciar la efectividad de diferentes procesos en sanidad, y como índice de productividad laboral y costes sanitarios. Por ejemplo, se estima que la mejora de una media de 10 puntos en la escala del SF-36 representa una disminución del 57% en la población general que refieren dolor de moderado a extremo en su mano al trabajar y que un cambio en el componente estado físico de 35 a 42 puntos predice un

33% de disminución en las bajas laborales. Asimismo, también se ha observado una mejoría en la salud mental después de la cirugía. (Atroshi 1999)

En el análisis de la sensibilidad al cambio del instrumento 6-CTS desarrollado por Atroshi (Atroshi 2011), se obtuvo una puntuación del efecto de tamaño de 2.0, similar al efecto tamaño de 2.1 de la escala de severidad de síntomas de 11 ítems. Además se obtuvo una MCID de 0.9. En nuestro estudio con 40 pacientes, se obtuvo un ES de 3.06 y una SRM de 2.94 del 6-CTS, lo cual apoya nuestra hipótesis de que el instrumento 6-CTS tiene una alta sensibilidad para detectar cambios clínicos tras tratamiento del STC, mayor que el QuickDASH con un ES de 1.62 y una SRM de 1.34 y mayor que el EQ-5D con un ES de 0.76 y una SRM de 0.79.

Estos resultados pueden ser muy útiles a la hora de demostrar el efecto del tratamiento en estudios sobre cirugía del STC, teniendo en cuenta que depende tanto de la muestra como del tipo de tratamiento, la MCID debe ser considerada dentro de un rango de valores y necesita futuras confirmaciones. Tradicionalmente, los tratamientos con un efecto tamaño de 0.5 son considerados moderadamente efectivos. Estos resultados deben ser considerados de manera específica según el tipo de tratamiento. Por ejemplo, en el tratamiento no quirúrgico del STC se obtuvo un ES de 0.5-0.7 con el 6-CTS, que si lo comparamos con el ES de la cirugía, se considera un pequeño efecto y no un moderado efecto. Cuando comparamos 2 tratamientos quirúrgicos que no se espera que tengan diferencia respecto al resultado, pero que potencialmente difieren en el grado de dolor postquirúrgico y limitaciones en la actividad, la escala del dolor palmar CTS, podría ser más adecuada como primer instrumento de medición.

Para el estudio de la valoración de resultados en investigación clínica existen diversos instrumentos que pueden ser elegidos. Es trabajo del investigador el de seleccionar aquél que se aproxime más a sus objetivos. Aunque esto puede parecer

evidente es un error bastante común la selección de un instrumento inadecuado. Lo mejor es siempre usar un instrumento genérico y uno específico.

Nuestro trabajo avala la hipótesis planteada: la versión española del instrumento 6-CTS tiene una buena validez de construcción, confiabilidad y alta sensibilidad en detectar cambios clínicos, por lo que debe ser considerado como instrumento de elección para valoración de resultados a los 3 meses tras cirugía del STC.



# 5

## CONCLUSIONES

- 1) Se confirma que el nivel de evidencia científica de los artículos de investigación clínica en las revistas de cirugía de la mano más relevantes es bajo y dependiente de la variable tipo de revista, basado en el análisis de dos observadores independientes con un alto grado de fiabilidad.
- 2) Los conocimientos en metodología de investigación y Medicina Basada en la Evidencia se postulan como una formación importante en todo profesional clínico tanto para la gestión de recursos como para una adecuada aplicación clínica de los conocimientos médicos, basados en la obtención de la información más relevante de la investigación clínica publicada.

- 3) Se confirma que los instrumentos PRO son útiles en la investigación clínica, permiten valorar diferentes dimensiones de la salud y característicamente presentan una mayor sensibilidad para detectar cambios de importancia clínica, comparados con las medidas de estructura-funcionamiento ( medidas clásicas de la exploración física) a la hora de valorar resultados tras tratamiento del STC. Consecuentemente, las medidas de los instrumentos PRO deberían ser las variables “outcomes” o resultados primarios en cualquier estudio de investigación clínica en STC.
  
- 4) Los instrumentos CTS y DASH en su versión española, son las medidas más sensibles para detectar cambios en una población de pacientes afectados de síndrome del túnel carpiano, sometidos a tratamiento quirúrgico, a las doce semanas tras tratamiento, comparado con las medidas convencionales de exploración física.
  
- 5) Se confirma que los instrumentos específicos son más sensibles para detectar cambios en la patología a estudio que los genéricos. En este sentido el instrumento CTS fue más sensible que el DASH y éste más que el SF 36.
  
- 6) La versión corta del cuestionario CTS (CTS-6) es capaz de reducir la carga al entrevistado mientras mantiene las mismas propiedades psicométricas en cuanto a consistencia interna, fiabilidad, validez de construcción y respuesta.
  
- 7) Se demostró que la sensibilidad al cambio de la versión española del instrumento 6-CTS es mayor que con el instrumento QuickDASH y que con el instrumento EQ-5D.

**6**

**ANEXOS**



## **Anexo A: Tabla 1**

**Niveles de Evidencia Científica según el tipo de estudio  
de Investigación Clínica Oxford Centre for  
Evidence-Based Medicine Levels of Evidence  
(May 2001).**



**Anexo A: Tabla 1: Niveles de Evidencia Científica según el tipo de estudio de Investigación Clínica. OxfordCentre for Evidence -Based Medicine Levels of Evidence (May 2001)**

Nivel	Tratamiento/Prevención/ Etiología/Complicaciones	Pronóstico	Diagnóstico	Diagnóstico Diferencial/Estudio de Prevalencia de síntomas	Análisis económico y de decisiones
1a	RS(con homogeneidad*) de ECA	RS(con homogeneidad*)de estudios de cohortes;GPC validada en diferentes poblaciones	RS(con homogeneidad*)de estudios diagnósticos con nivel 1:GPC+ con 1b estudios de diferentes centros clínicos	RS(con homogeneidad*)de estudios de cohortes prospectivos	RS(con homogeneidad*)de estudios económicos con nivel1.
1b	ECA individual(con intervalo de confianza estrecho #)	Estudios de Cohortes Individuales con> = 80%de seguimiento; GPC+ validada en una única población	Estudios de cohorte que validen**la calidad de una prueba específica, con unos buenos estándares de referencia+++;o de una GPC probada en un centro clínico	Estudio de cohorte prospectivo con buen seguimiento****	Análisis basados en los costes clínicos o en sus alternativas: revisiones sistemáticas de la evidencia e inclusión de análisis de sensibilidad
1c	Todo o ninguno &	Todos o Ninguno series de Casos	Absoluta especificidad y sensibilidad.++	Todos o Ninguno serie de casos	Análisis del mejor valor o del peor valor absoluto ++++
2a	RS(con homogeneidad*) de estudios de cohortes	RS(con homogeneidad*) de estudios de cohortes retrospectivo o de grupos controles no tratados en un ECA.	RS(con homogeneidad*)de estudios diagnósticos de nivel >2.	RS(con homogeneidad*) de 2b y estudios mejores	RS(con homogeneidad*) de estudios económicos de Nivel>2.
2b	Estudio de cohorte individual (incluyendo ECA de baja calidad ;ej <80% de seguimiento).	Estudio de cohorte retrospectivo o seguimiento de controles no tratados en un ECA: derivación de GPC.	Estudios de cohortes exploratorios**con buenos estándares de referencia +++;o a partir de algoritmos de estimación del pronóstico o de categorización del diagnóstico, o de validación de muestras separadas&&&	Estudio de cohorte retrospectivo, o pobre seguimiento	Análisis basados en los costes clínicos o en sus alternativas; revisiones sistemáticas con evidencia limitada, o estudios individuales; e inclusión de análisis de sensibilidad
2c	Investigación de resultados en salud; estudios ecológicos	Investigación de resultados en salud		Estudios ecológicos	Investigación de resultados en salud
3a	RS(con homogeneidad*) de estudios de casos y controles		RS(con homogeneidad*) de 3b y estudios mejores	RS(con homogeneidad*) de 3b y estudios mejores	RS(con homogeneidad*) de 3b y estudios mejores
3b	Estudios de casos- controles individuales		Estudio no consecutivo: o sin referencias estándares consistentemente aplicadas	Estudios de cohorte no consecutivos, o población muy limitada	Análisis sin medidas de coste precisas pero incluyendo un análisis de sensibilidad que incorpora variaciones clínicamente sensibles en las variables importantes
4	Serie de casos(estudio de cohortes y casos y controles de baja calidad &&)	Serie de casos (estudio de cohortes de pronóstico de poca calidad **)	Estudio caso-control, con pobre o no independiente estándar de referencia.	Serie de casos o estándar de referencia sustituida	Análisis que no incluye análisis de la sensibilidad
5	Opinión de expertos sin valoración crítica explícita, ni basada en fisiología, ni en investigación juiciosa ni en los principios fundamentales	Opinión de expertos sin valoración crítica explícita, ni basada en fisiología, ni en investigación juiciosa ni en los principios fundamentales	Opinión de expertos sin valoración crítica explícita, ni basada en fisiología, ni en investigación juiciosa ni en los principios fundamentales	Opinión de expertos sin valoración crítica explícita, ni basada en fisiología, ni en investigación juiciosa ni en los principios fundamentales	Opinión de expertos sin valoración crítica explícita, ni basada en teorías económicas ni en los principios fundamentales

RS (revisión sistemática); ECA (ensayos clínicos aleatorizados)

\* Con homogeneidad, es decir, que incluya estudios con resultados comparables y en la misma dirección y GPC (guía práctica clínica) validadas. Libre de variaciones erróneas (heterogeneidad). No todas las revisiones sistemáticas con heterogeneidad estadísticamente significativas son erróneas, y no toda heterogeneidad errónea es estadísticamente significativa. Estudios heterogéneos deben ser señalados con un signo negativo al final de su nivel designado."-".

+ GPC: Guía Práctica Clínica. Son algoritmos o sistemas de puntuación que conducen a una estimación pronóstica o a una categoría diagnóstica.

# aplicar amplios intervalos de confianza.

& Cuando todos los pacientes mueren antes de que un determinado tratamiento esté disponible, y con él algunos pacientes sobreviven, o bien cuando algunos pacientes morirían antes de su disponibilidad, y con él no muere ninguno.

&& Por estudio de cohorte de pobre calidad entendemos estudios sin clara definición de los grupos comparados y/o sin medición objetiva de las exposiciones y eventos (preferentemente ciega) y/o sin identificar o controlar adecuadamente variables de confusión conocidas y/o sin seguimiento completo y/o suficientemente prolongado. Por estudio de caso-control de pobre calidad, entendemos estudios sin clara definición de los grupos comparados y/o sin medición objetiva de las exposiciones y eventos (preferentemente ciega) y/o sin identificar o controlar adecuadamente variables de confusión conocidas.

---

&&& "Split- sample validation" se obtiene al recolectar todos los datos de una vez, luego se divide artificialmente en "muestra de derivación" y "muestra de validación".

++ Pruebas diagnósticas con especificidad tan alta que un resultado positivo confirma el diagnóstico y con sensibilidad tan alta que un resultado negativo descarta el diagnóstico.

## Bueno, mejor, malo y peor se refiere a comparaciones entre tratamientos en términos de sus riesgos y beneficios clínicos.

+++ Los estándares de referencia buenos son independientes de la prueba, y son aplicados a ciegas y de forma objetiva a todos los pacientes. Los estándares de referencia pobres son aplicados fortuitamente, pero independientes de la prueba. El uso de un estándar de referencia no independiente (donde el test es incluido en la referencia, o donde las preguntas afectan a la referencia) implica un estudio de nivel 4.

++++ Tratamientos mejores son claramente tan buenos pero más baratos, o mejores o iguales y coste reducido. Tratamientos peores son tan buenos y más caros, o peores o iguales y más caros \*\* Los estudios que validan la calidad de una prueba específica, se basan en la evidencia fundamental. Un estudio exploratorio recoge información y arrastra los datos (ej: usando un análisis de regresión) para encontrar qué factores son significativos.

\*\*\* Por estudios de cohorte de escasa calidad entendemos estudios en los que el muestreo fue con prejuicios, a favor de pacientes que ya tenían seguimiento, o el seguimiento fue < 80% de los pacientes del estudio, o sin medición objetiva o sin controlar variables de confusión conocidas.

\*\*\*\* Un estudio de diagnóstico diferencial bueno tiene un seguimiento de > 80%, con un tiempo adecuado para nuevos diagnósticos alternativos (ej agudo 1-6 meses, crónico 1-5 años).

## **Anexo B: artículo original**

### **“Level of evidence in hand surgery”**

Roberto S Rosales <sup>1</sup> . Luis Rebozo-Morales <sup>2</sup>. Yolanda Martín Hidalgo <sup>2</sup>.

Isabel Díaz de la Lastra-Bosch <sup>3</sup>

<sup>1</sup>Corresponding author Email: titorosales telefonica.net. Unit for Hand & Microsurgery, GECOT, María del Cristo Ossuna 20, La Laguna, Tenerife, 38204, Spain

<sup>2</sup>Orthopaedic Surgery Service, University Hospital of "La Candelaria", Carretera del Rosario s/n, La Laguna, Tenerife 38110, Spain

<sup>3</sup>Government Health Service, Oncology Programme, Canarian Health Service, Pérez de Rozas 5, 2nd floor, Santa Cruz de Tenerife, Tenerife 38005, Spain.

*BMC Res Notes. 2012;5:665*

## **Abstract**

### **Background**

Few investigations have been done to analyze the level of evidence in journals related to hand surgery, compared to other related research fields. The objective of this study was to assess the level of evidence of the clinical research papers published in the Ibero-American (RICMA), the European (JHSE) and American (JHSA) Journals of Hand Surgery.

### **Methods**

A total of 932 clinical research papers published between 2005 and 2009 (RICMA 60, JHSE 461, and JHSA 411) were reviewed. Two independent observers classified the level of evidence based on the Oxford International Classification, 5 being the lowest level and 1 the highest level. The observed frequencies of the level of evidence for each journal were compared with the expected frequencies by a chi-square ( $\chi^2$ ) test for categorical variables with a significance level of 0.05.

### **Results**

Inter-observer agreement analysis showed a Kappa of 0.617. Intra-observer agreement analysis presented a kappa of 0.66 for the observer 1, and a Kappa of 0.751 for the observer 2. More than 80% of the papers in RICMA and JHSE and 67.6% in the JHSA presented a level of 4. No level 1 or 2 studies were published in RICMA, compared to JHSE (0.9% level 1 and 5.0% level 2) and JHSA (8.3% level 1 and 10% level 2). The percentage of papers with level 3 published in RICMA

(16.7%) was higher compared to the JHSE (11.1%) and JHSA (14.1%). All the results were statistically significant ( $x = 63.945$ ;  $p < 0.001$ ).

## **Conclusions**

The level of evidence in hand surgery is dependent on the type of journal; being the highest level evidence papers those published in the JHSA, followed by the JHSE and finally the RICMA. Knowing the status of the level of evidence published in hand surgery is the starting point to face the challenges of improving the quality of our clinical research.

**Keywords:** Hand, Level of evidence, Hand Surgery, Clinical Research

## **Background**

Since the first system for classifying the level of evidence of the clinical research papers was reported (1), Evidence- Based Medicine has become an important part of our clinical practice. Hand surgeons should understand the level of evidence in order to become aware of the reliability and the utility of the data provided in a research paper. Few investigations have been done to analyze the level of evidence in journals related to hand surgery compared to other related research fields, as in the orthopaedic surgery (2) (3), and plastic surgery (4) journals. Only one specific hand surgery journal has been analyzed for evidence level over a six month period of time, and has been compared to other orthopaedic publications (3). To our knowledge, no paper comparing the level of evidence in clinical research published in three hand surgery specific journals over a period of five years, has been reported before. The purpose of this paper was to assess the level of evidence of the clinical research papers published between 2005 and 2009 in the Ibero-American Journal of Hand Surgery (RICMA), as the official journal of the Spanish, Portuguese and the main

Latin American Societies for Surgery of the Hand; the European (JHSE) and American (JHSA)

Journals of Hand Surgery, as the official journals of the Federation of European Societies for Surgery of the Hand (FESSH) and the American Society for Surgery of the Hand (ASSH).

## **Hypothesis**

The researches established the null hypothesis (Ho) that the variable "level of scientific evidence" was independent of the variable "type of journal".

## **Methods**

### **Elegibility criteria and population study**

Inclusion criteria. All the clinical research articles, which were published between January 2005 and December 2009 in the: Ibero- American of Hand Surgery (RICMA) ("Revista Iberoamericana de Cirugía de la Mano "); The Journal of Hand Surgery European Volume (JHSE) and The Journal of Hand Surgery American Volume (JHSA).

Exclusion criteria: Animal studies, anatomical of cadaver studies, basic science studies, instructional course lectures, supplements of abstract, short reports, letters to the editors and review articles were not considered as feasible for the study.

Hence, a total of 932 clinical research papers followed the inclusion and exclusion criteria (RICMA 60, JHSE 461, and JHS 411).

---

## **Assessment of level of evidence**

The articles which met the inclusion and exclusion criteria were randomly assigned to two independent observers (YMH, LRM), with large clinical experience in hand surgery, and very familiar with Evidence-Based Medicine. An approximate equal number of articles from each journal were assessed based on The Oxford Centre for Evidence -Based Medicine- Levels of Evidence (March 2009) <http://www.cebm.net/index.aspx> (5). The observers were blinded to any previous information related to the level of evidence of the articles to be analysed (e.g.: "level of evidence and type of study", which is included in the JHSA with the abstract of every clinical research paper since 2006). The articles were ranked according to their level of evidence from Level 1 (highest level of evidence e.g.: systematic review (SR), meta-analysis (MA), and individual randomized controlled trials (RCT) with narrow interval confidence) to level 5 (lowest level of evidence, e.g.: expert opinion). Assessment of the sub-group level of evidence, as well as the type of research article was not done by the observers (Table 1).

Table 1 Level of evidence and type of study

	Therapy /Prevention, Aetiology/Harm	Prognosis	Diagnosis	Differential Diagnosis/ symptom prevalence	Economic and decision analysis
		Investigating the effect of patient characteristic on the outcome of the disease	Investigating a diagnostic test. Is this diagnostic test accurate?		
<b>Level 1</b>	Systematic Review of randomized trials(RT)  High quality RT(e.g.:> 80% follow up, narrow confident interval)	Systematic Review of inception cohort studies  Individual cohort study with > 80% follow up, all patient enrolled at the same time	Systematic Review of level 1 diagnostic studies  Level 1 diagnostic studies or Validating studies which test the quality of a specific diagnostic test, previously developed, in series of consecutive patients with reference "gold" standard	Systematic Review of prospective or classic cohort  Prospective or classic cohort studies with good follow up (>80%)	Systematic Review of level 1 economic studies  Level 1 studies (analysis based on clinically sensible costs or alternative, values obtained from many studies, and including multiway sensitive analysis)
<b>Level 2</b>	Systematic Review of cohort studies  Lesser quality RT (e.g.: <80% follow up, wide confident interval, no clear randomization, problems with blinding, etc.)  Individual Cohort study, including matched cohort studies (prospective comparative studies)  Ecological Studies	Systematic Review of either historical cohort study or untreated control groups (control arm) in RCTs  Historical (retrospective) cohort study or control arm from a RCT	Systematic Review of level 2 diagnostic studies  Level 2 diagnostic studies or Exploratory studies which collect information, trawl data to find which factor are significant (e.g.: using regression analysis)	Systematic Review of level 2 studies  Level 2 studies (retrospective or historical cohort study or with follow up <80%)  Ecological Studies	Systematic Review of level 2 studies  Level 2 studies (analysis based on clinically sensible cost or alternative from limited studies, and including multiway sensitivity analysis.
<b>Level 3</b>	Systematic Review of case-control studies  Individual case-control study		Systematic Review of level 3 studies  Level 3 diagnostic studies or studies in non-consecutive patients and without consistently reference "gold" standards	Systematic Review of level 3 studies  Level 3 studies (non-consecutive cohort or very limited population)	Systematic Review of level 3 studies  Level 3 studies (analysis based on poor alternative or costs, poor quality estimates of data, but including sensitivity analysis)
<b>Level 4</b>	Case-series  Poor quality cohort and case-control studies*	Case-series  Poor quality cohort and case-control studies*	Case-control study  Poor or non independent reference standard	Case-series	No sensitivity analysis
<b>Level 5</b>	Expert opinion	Expert opinion	Expert opinion	Expert opinion	Expert opinion

A systematic review (SR) is generally better than an individual study. Experimental study (e.g.: good quality RCT) is generally better than any observational study. For observational studies: cohort study is generally better than any case–control study. A case- control study is generally better than any case- series study.

\* By poor quality cohort study we mean a cohort study that failed to clearly define comparison groups and/or failed to measure exposures and outcomes (preferable blinding) in the same objective way in both expose and non-exposed individuals and/or failed to identify control known confounders and/or poor follow up.

The same for poor quality case–control study except that the patients are identified based on the outcomes in this design (e.g.: failed replant) called “cases” are compared with those who did not have the outcome (e.g.: had a successful replant) called “controls” and consequently we do not have “exposed and non-exposed” and “longitudinal follow up”. Ecological studies and

Economic/decision analysis studies are very uncommon in hand surgery. This chart was adapted from material published by the Centre for Evidence Based medicine, Oxford, UK March 2009.

### **Reliability analysis**

Before starting the study, the reliability of the assessment was evaluated based on the analysis of both the intra-observer error and inter-observer error. A random sample of 30 clinical research articles, from a total of 872 papers, published in the English language (461 from JHSE, and 411 from JHSA), were assessed by the two independent observers assigned to the study. After 15 days, a second assessment was

undertaken with the order of the articles changed. No papers from the RICMA were included in the sample study for the reliability analysis. This was done so to avoid information bias (6), because the different languages present in the RICMA publication (Spanish and Portuguese), could increase the intra - observer reliability. The intra-observer and inter-observer reliability was studied using the Kappa coefficient with a significance level of 0.05.

### **Data analysis**

For the assessment of the results, the number of articles for each level of evidence rating was expressed as a percentage of the total number of articles meeting the inclusion and exclusion criteria for the period time study. The observed frequencies of the level of evidence for each journal were compared with the expected frequencies using a chi-square ( $\chi^2$ ) test for categorical variables with a significance level of 0.05.

### **Results**

Inter-observer analysis showed Kappa of 0.617, with an asymptotic standard error of Kappa (SE) of 0.117. Intra-observer analysis presented a Kappa of 0.66 (SE 0.114) for observer 1, and a Kappa of 0.751 (SE 0.103) for observer 2. All Kappa values were significant ( $p < 0.001$ ). More than 80% of the papers in the RICMA and JHSE; and a 67.6% in the JHSA presented a level 4. Not one of the level 1 and 2 papers was published in the RICMA compared to the JHSE (0.9% level 1 and 5.0% level 2) and the JHSA (8.3% level 1 and 10% level 2). The percentage of papers with level 3, published in the RICMA (16.7%) was higher compared to the JHSE (11.1%) and the JHSA (14.1%) (Table 2). All the results were statistically significant ( $\chi^2 = 63.945$ ;  $p < 0.001$ ) and the null Hypothesis ( $H_0$ ) was rejected.

Cross tabulation of "type of journal" and "level of evidence"

Journal	Level of Evidence					Nºpapers
	1	2	3	4	5	
RICMA	0.0%	0.0%	16.7%	80.0%	3.3%	60
(CI 95%)	(N.A)	(N.A)	(7.3;26.1)	(69.9;990.1)		
	0	0	10	48	2	
JHSE	0.9%	5.0%	11.1%	82.4%	0.7%	461
(CI 95%)	(N.A)	(3.02;6.9)	(8.3;13.9)	(79;85.8)	(N.A.)	
	4	23	51	380	3	
JHSA	8.3%	10%	14.1%	67.6%	0%	411
(CI 95%)	(5.6;10.9)	(7.1;12.9)	(10.7;17.4)	(63.1;72.1)	(N.A.)	
	34	41	58	278	0	

RICMA = Ibero-American Journal of Hand Surgery, JHSE= Journal of Hand Surgery European Volume, JHSA= Journal of Hand Surgery American Volume. (CI 95%) =95% Confidence Interval. (N.A.) = CI 95% is not applicable when the observed proportion is not greater than 5/n.

## Discussion

Results of this paper have demonstrated with a good-excellent level of reliability that the variable "level of evidence" is dependent on the variable "type of journal".

## **Reliability analysis**

The use of Kappa is important, as an often used proportion of agreement does not allow for the fact that some agreement is due to chance. A statistically significant Kappa coefficient means that the agreement is different from zero (null agreement).

However, the interpretation of obtained values of Kappa is subjective, and different classifications or guides have been proposed to interpret the kappa coefficient in the reliability analysis. In this paper, the level of agreement in the inter- observer and intra - observer analysis has shown that a kappa value ranging from 0.617 to 0.751, can be considered as having an excellent to a good level of reliability (7,8) in the assessment of the level of evidence and the type of journal. Similar results have been reported before. Obremskey et al. (3), in the assessment of the level of evidence in orthopaedics journals, have reported Kappa values of 0.62 for inter-observer agreement between inexperienced reviewers, and a kappa value of 0.75 for inter-observer reliability between experienced reviewers. No intra observer agreement analysis was reported by those authors.

## **Level of evidence and type of journal**

Not many papers have studied the level of evidence in hand surgery journals or in related research fields, such as orthopedics and plastic surgery journals. Sinno et al. (4), reviewed 726 from six different plastic surgery journals and the level of evidence was assessed using a classification based on the Oxford Centre for Evidence level (CEBM). Hanzlik et al. (12) assessed 551 papers from the Journal of Bone Joint Surgery American Volume (JBJS-A) from the years 1975 (134 papers), 1985 (123 papers), 1995 (120 papers), and 2005 (174 papers). The level of evidence was

assessed using a classification included in the guide for authors (JBJS-A grading system) which was very similar to the one developed by the CEBM, in order to demonstrate trends in the level of evidence over 30 years. Furthermore, Obremskey et al. [3] reviewed 382 clinical research articles from nine different journals in order to assess the level of evidence in orthopaedic journals. In this paper, 932 clinical research papers from three specific hand surgery journals were reviewed, which constitutes the largest population of scientific clinical articles assessed to study the level of evidence reported until now.

The results of this paper demonstrate that most of the clinical articles published in hand surgery, are papers with a very low level of evidence (80% level 4 in the JHSE or RICMA and 67.6% in the JHSA). Most of those papers were case-series and less frequently, poor quality cohort or poor quality case-control studies. Those results were higher compared to orthopedic journals (48 % level 4 studies) [2], to plastic surgery journals (40% level 4 studies) [4] and to ophthalmology journals (58% Level 4 studies) [9]. However, other surgical journals as ear, nose and throat (otolaryngology) journals present a percentage similar to JHSE and RICMA (80% Level 4 studies) [10].

The percentage of level 4 papers in JHSA was lower, as compared to the rest of the hand surgery journals investigated, and it was very close to the one published by Obremskey et al. [3], who reported a 68.8% of level 4 papers, in a review of 32 articles published in the JHSA from January to June 2003.

The percentage of papers with a higher level of evidence (level 1 and 2), was larger in the JHSA (8.3% level 1 and 10% level 2), compared to the RICMA (0%) and the JHSE (0.9% level 1 and 5% level 2). Whilst compared to other journals, there was 21% of level 1 and 15% of level 2 of evidence in orthopaedic journals [2], 3% of

level 1 and 16% of level 2 in plastic surgery journals [4], 18% of level 1 and 8% level 2 in ophthalmology journals [9], and 7% of level 1 and level 2 in otolaryngology journals [10].

The percentage of papers with level 3 (mostly case-control studies and non-consecutive cohort studies or with very limited population) published in the RICMA (16.7%) was higher compared to the JHSE (11.1%) and the JHSA (14.1%); and similar to other journals: 16% in orthopaedic journals [2], 16% in otolaryngology journals [10] and 16% in ophthalmology journals [9]. Hence, some authors have criticized the low number of high evidence level in surgery [11]. Even so, the criticism may seem overly severe, if we take into account that surgical trials are different from trials, which compare a medication with a placebo. Surgical procedures are invasive; it is difficult to randomise patients, blinding is a problem in surgical trials, and they are very expensive. If we do not have high quality randomized trials we cannot have a systematic review which synthesizes the evidence previously reported. No trend analysis is a limitation for this paper, and the information within should be the purpose of further studies, in order to understand how the evidence published in hand surgery journals has changed and how the relationship between changes in the level of evidence and changes in the impact factor index, have also changed over time.

After reviewing several articles published in journals from different parts of the world, other questions have arisen. These being, whether the differences that we have found are a reflection of different regional priorities or how the resources used for research have an impact on our findings and even if different countries are the main contributors in high level studies.

## **Conclusions**

The level of evidence in hand surgery is dependent on the type of journal; being the highest level evidence papers those published in the JHSA, followed by the JHSE and finally the RICMA. Knowing the status of the level of evidence published in hand surgery is the starting point to face the challenges of improving the quality of our clinical research.

## **Competing interests**

There are no financial or non-financial competing interests to declare in relation to this manuscript.

## **Authors' contributions**

RSR was involved as director of the study. He devised and designed the study and was also involved in the analysis and the interpretation of the data. LRB & YMH were the two independent observers involved in the reliability analysis and in the assessment of the level of evidence of papers which followed the inclusion and exclusion criteria. IDLB was involved in the acquisition and analysis of the data, as well as in the drafting and reviewing of the manuscript. All the authors read and approved the final version of the manuscript.

## **Acknowledgement**

The authors of this paper thank Mrs. Estefania García Mesa, Prof. of English language for her contribution to this paper.

## **Author details**

<sup>1</sup> Unit for Hand & Microsurgery, GECOT, Maria del Cristo Ossuna 20, La Laguna, Tenerife, 38204, Spain.

<sup>2</sup>Orthopaedic Surgery Service, University Hospital of “La Candelaria”, Carretera del Rosario s/n, La Laguna, Tenerife 38110, Spain.

<sup>3</sup>Government Health Service, Oncology Programme, Canarian Health Service, Pérez de Rozas 5, 2nd floor, Santa Cruz de Tenerife, Tenerife , 38005, Spain.

## References

1. Sackett DL: Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest* 1986, 89 (2 Suppl):2S–3S.
2. Hanzlik S, Mohair RC, Baynosa RC, Khiabani KT: Levels of evidence in research published in the journal of bone and joint surgery (American volume) over the last thirty years. *J Bone Joint Surg* 2009, 91 A:425–428.
3. Obremskey WT, Pappas N, Attallah-Wasif E, Tornetta P, Bhandari M: Level of evidence in orthopedic journals. *J Bone Joint Surg* 2005, 87 A:2632–2638.
4. Sinno H, Neel OF, Lutfy J, Bartlett G, Gilardino M: Level of evidence in plastic surgery research. *Plast Reconstr Surg* 2011, 127:974–980.
5. Oxford Centre for Evidence-based Medicine: Levels of Evidence. Available at: <http://www.cebm.net/index.aspx?o=1025>. Accessed March 2009.
6. Page RM, Cole GE, Timmreck TC: *Basic Epidemiological Methods and Biostatistics. A practical guidebook*. Boston: Jones and Bartlett publishers; 1995.
7. Landis JR, Koch GG: The measurements of observer agreement for categorical data. *Biometrics* 1977, 33:159–174.
8. Silman AJ: *Epidemiological Studies: a practical guide*. New York: Cambridge University Press; 1995.
9. Lai TY, Leung GM, Wong VW, Lam RF, Cheng AC, Lam DS: How evidence-based are publications in clinical ophthalmic journals? *Invest Ophthalmol Vis Sci* 2006, 47:1831–1838.
10. Bentsianov BL, Boruk M, Rosendfield RM: Evidence-based medicine in otolaryngology journals. *Otolaryngol Head Neck Surg* 2002, 126:371–376.
11. Horton R: Surgical research or comic opera: questions, but few answers. *Lancet* 1996, 347:984–985.



## **Anexo C: artículo original**

# **“The Relative Responsiveness and Construct Validity of the Spanish Version of the DASH Instrument for Outcomes Assessment in Open Carpal Tunnel Release”**

ROBERTO SÁNCHEZ ROSALES<sup>1</sup>, ISABEL DIEZ DE LA LASTRA<sup>2</sup>,  
STEVEN.MCCABE<sup>3</sup>, JOSE IGNACIO ORTEGA MARTINEZ<sup>4</sup> and Y.  
M. HIDALGO<sup>5</sup>

<sup>1</sup>Unit for hand & Microsurgery, GECOT, María del Cristo Ossuna 20, La Laguna, Tenerife, 38204, Tenerife.

<sup>2</sup>Department of Bioinformatics and Biostatistics, School of public Health and Information Sciences

<sup>3</sup>University of Louisville, Louisville, KY, USA

<sup>4</sup>Private Practice & Plastic Surgery Service, Hospital Virgen del Rocío, Sevilla, Spain

<sup>5</sup>Orthopaedic Surgery Service, University Hospital of "La Candelaria", Carretera del Rosario s/n, La Laguna, Tenerife 38110, Spain

---

*J Hand Surg Eur Vol 2009 34: 72 originally published online 17 December 2008.*

### **Abstract**

The objective of this study was to evaluate the responsiveness and construct validity of the Spanish version of the DASH as outcome measure for carpal tunnel surgery. The study population was 42 patients with the diagnosis of carpal tunnel syndrome (CTS) based on clinical and electrophysiological criteria. The clinical design was a classic Cohort study with measures the day before and 12 weeks after open carpal tunnel release. The Spanish version of the DASH was compared to the physical exam measures as dexterity time, grip and pinch strength, range of motion of the wrist (ROM) and two-point discrimination (2PPD). The responsiveness was evaluated based on the effect size (ES) and the standardised response means (SRM). The Spanish version of the CTS questionnaire (CTQ) and the Spanish version of the 36-item short-form health survey (SF-36) were self-administered to the same study population and followed the same clinical design. The hypothesis that the DASH instrument should present a responsiveness level higher than the SF-36 and lower than the CTQ was established to demonstrate the construct validity. The DASH instrument showed an ES of 0.68 and an SRM of 1.00. Responsiveness of traditional physical exam measures were lower, running from 0.35 (SRM) for dexterity time to 0.00 (ES) for key pinch strength. The SF-36 presented a responsiveness level (range from 0.07 to 0.14) lower than the DASH. The CTQ showed the highest level of responsiveness (ES= 1.41 and 0.7) (SRM = 1.75 and 0.51). In conclusion, The DASH instrument is more sensitive in detecting clinical change than the physical exam measures for use in clinical outcome studies of CTS done at 12 weeks after surgery. The Spanish DASH showed

responsiveness lower than the CTQ and higher than the SF-36 as a proof of good construct validity.

**Keywords:** outcomes, responsiveness, validity, DASH, carpal tunnel syndrome

## **Introduction**

The cross-cultural adaptation of a health instrument is a long process that does not finish with the translation– adaptation. It should ideally include all psychometric analyses (reliability, validity and responsiveness). The cross-cultural adaptation of the DASH to Spanish and the evaluation of the Spanish version’s reliability in patients with carpal tunnel syndrome (CTS) have been reported; however, responsiveness and construct validity evaluation were not included in the study (Rosales et al., 2002).

Responsiveness is the ability of an instrument to measure a clinically meaningful or important clinical change in a clinical stage (Liang, 1995). The DASH has been shown to be comparable with the CTS questionnaire (CTQ) for both wrist and hand conditions in general (Beaton et al., 2001). It has also been shown that the DASH is sufficiently responsive for its use in outcome studies of CTS and that its sensitivity to detect important clinical change (at 12 weeks or more after treatment) is intermediate compared to the CTQ (Levine et al., 1993) and the 36-item short-form health survey (SF-36) instruments. These results constituted evidence of the construct validity of the DASH instrument for outcome assessment in CTS (Gay et al., 2003). The purpose of this paper is to demonstrate the construct validity of the Spanish version of the DASH, and its relative responsiveness compared to the traditional physical exam measurements. The hypothesis to be tested was: The Spanish DASH applied to a

specific CTS population will present a level of responsiveness higher than that of a generic instrument such as SF-36, and lower than that of a specific disease instrument as in the CTQ.

## **Material and Methods**

### **Study Population**

The institutional review board of our medical centre approved the protocol. Informed consent was obtained from each subject before participation. Forty-two consecutive patients were enrolled in this study with a diagnosis of CTS based on clinical and electrophysiological criteria. Inclusion criteria were: (1) numbness or tingling with or without pain in at least two of the digits of the median nerve distribution (Atroshi et al., 1999; Gay et al., 2003); (2) increased symptoms with carpal tunnel provocation tests (Phalen's test, and/or reverse Phalen) (Gay et al., 2003); (3) symptoms for over 2 months (Atroshi et al., 1999); (4) failure in conservative treatment (Atroshi et al., 1999) and (5) positive electrophysiological testing (Amadio et al., 1996; Atroshi et al., 1999; Gay et al., 2003). Exclusion criteria were clinical or electrophysiological signs of proximal nerve compression, diabetes or any metabolic disease, rheumatoid arthritis or other general inflammatory diseases (Amadio et al., 1996; Atroshi et al., 1999; Gay et al., 2003).

### **Clinical design**

This was a cohort study from the day before the operation to 12 weeks after open carpal tunnel release under Brachial Plexus block. One surgeon performed all the operations.

## **Instruments and measures**

An independent and fully trained hand surgeon evaluated all the patients the day before surgery and 12 weeks after. The physical examinations were dexterity time, two-point discrimination (2PPD), grip strength, key pinch, pulp to pulp pinch and range of motion of the wrist (ROM). The assessment followed a standardised protocol: dexterity was measured by a timed pick-up test. Patients were asked to pick up five coins from the table, recording the mean time of three trials. 2 PPD was measured on the radial and ulnar aspects of the pulp of each digit. 2 PPD was recorded on 5 to 15 mm scale. The mean value for the three radial digits was used in the analysis. Grip strength was measured with the Jamar dynamometer (Preston, Jackson, MI) and key pinch and pulp-to-pulp pinch with the pinch gauge (B&L, Santa Fe Springs, CA), recording the mean of three trials. ROM was measured by a regular goniometer, recording the total active ROM (flexion + extension + ulnar deviation, radial deviation). In order to demonstrate the construct validity hypothesis, the Spanish version of the CTS instrument (Rosales et al., 2002) and the standard Spanish version of the medical outcomes study SF-36 (Alonso et al., 1995) were self-administered to the same population under study and following the same clinical design with measurements taken the day before and 12 weeks after surgery.

## **Data analysis**

Responsiveness for each measure was evaluated by calculating the effect size (ES) (mean change/standard deviation (SD) of the baseline value) and the standardised response mean (SRM) (mean change/SD of change). For the construct validity analysis, the responsiveness of DASH was compared to the responsiveness of the

functional status scale and symptoms severity scale of the CTQ, and to the role physical subscale (SF-36rp) and bodily pain subscale (SF-36bp) of the SF-36 instrument.

## Results

All 42 patients completed the study. There were 36 women and six men with a mean age of 54 (range 34–63) years. The mean and SD of the pre-operative and postoperative results and of the difference between the pre-operative and postoperative results for each measurement are listed in Table 1. The responsiveness for each measure is shown so that a positive number means improvement, and a negative value means deterioration in outcome. The DASH questionnaire presented an ES of 0.68 and an SRM of 1.00. The physical examination measures were less responsive, running from 0.35 (SRM) and 0.30 (ES) for dexterity time to 0.00 (SRM) (ES) for key pinch strength. Grip strength and wrist ROM gave negative values (Table 1). The responsiveness of the SF-36 measures were 0.07 (ES) and 0.08 (SRM) for SF-36bp and 0.14 (ES) and 0.12 (SRM) for SF-36rp. The CTQ had the highest responsiveness (range 0.51–1.41) (Table 1, Fig 1).

Tabla 1: responsiveness del DASH, CTQ, SF-36 and physical exam measures

Measure	Pre-op mean (SD)	Postop mean (SD)	Pre-post mean (SD)	E.S.	SR.M.
CTQ-SS	3.26 (0.92)	1.96 (0.68)	1.30 (0.74)	1.41	1.75
CTQ-EF	2.94 (0.84)	2.35 (0.71)	0.59 (1.14)	0.7	0.51
DASH	51.01 (19.62)	37.63 (51.01)	13.37 (13.25)	0.68	1.00
SF-36rp	20.31 (33.19)	25 (44.72)	4.69 (38.96)	0.14	0.12
SF-36bp	42.44 (22.63)	44.06 (21.59)	1.63 (18.87)	0.07	0.08
Dexterity time	5.94 (3.56)	4.48 (1.44)	1.06 (3.06)	0.30	0.35
Grip (kg)	26 (7.46)	24.88 (7.41)	-1.12 (4.51)	-0.15	-0.25
P-P pinch (kg)	5.71 (1.49)	5.97 (1.54)	0.26 (1.16)	0.17	0.22
Key pinch (kg)	6.35 (1.84)	6.35 (1.21)	0.00	0.00	0.00
ROM	203.53 (27.43)	195.53 (19.91)	-8.00 (15.26)	-0.29	-0.52
2PPD (mm)	5.53 (1.18)	5.35 (1.00)	0.18 (1.29)	0.18	0.14

The scores for the DASH range from 0 (no disability) to 100 (most severe disability); for the SF-36 range from 0 (worst health) to 100 (best health); for the CTQ higher scores mean worse outcomes. Pre-op = measurement before surgery; postop = measurement at 12 weeks after surgery, pre–post = change in scores from before to 12 weeks after surgery, SD =standard deviation, E.S = effect size, S.R.M = standardised response mean.

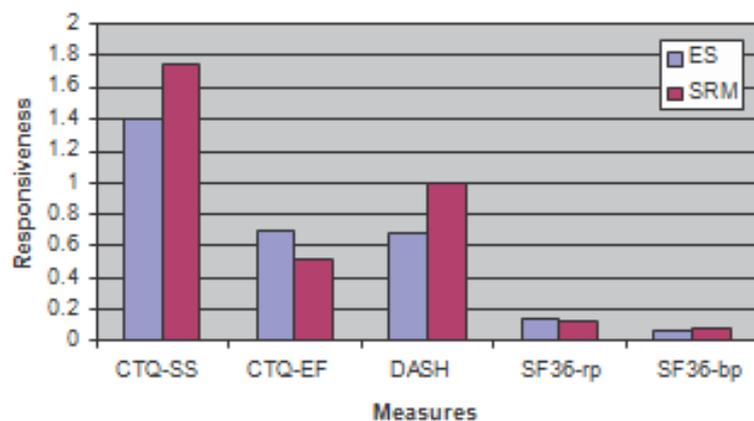


Fig 1 Construct validity of the DASH. The responsiveness of the DASH (upper extremities specific instrument) is higher than SF-36 (generic instrument) and lower than CTQ (disease-specific instrument). ES =effect size; SRM = standardised response mean.

## Discussion

The results show that the Spanish DASH is more responsive than physical examination and the SF-36rp and SF-36bp (generic instrument) but less than the CTQ (disease-specific instrument) for the treatment of CTS.

## **Cross-cultural equivalence and DASH responsiveness**

Several standard versions of the DASH have not reported responsiveness analysis in the original publications (Lee et al., 2005; Orfale et al., 2005; Padua et al., 2003; Rosales et al., 2002; Themistocleus et al., 2006). The Spanish DASH instrument presented a good responsiveness level at 12 weeks after carpal tunnel release with an ES of 0.68 and an SRM of 1.00. A similar level of responsiveness has been reported by other authors: Kotsis and Chung (2005) have reported an SRM with a medium effect of 0.7 (American version); De Smet et al. (2007) (Dutch version) presented an ES of 0.87, and an SRM of 0.69; Greenslade et al. (2004) (British population) showed an SRM of 0.62; and Gay et al. (2003) (American version) have shown an ES of 1.01, and an SRM of 1.13. Gummenson et al. (2003) with the Swedish version demonstrated at 6 months after surgery an ES of 0.7, and an SRM of 1.0. A lower rate of responsiveness of the DASH (ES of 0.49, SRM of 0.43) has been reported in a British population of CTS by Hobby et al. (2005); however, a high drop-out rate (25%) in the cohort population makes those results unsuitable for cross-cultural comparison. The Spanish version of the DASH detects a medium-to-large meaningful clinical change after carpal tunnel surgery similar to other standard and equivalent versions.

## **Treatment effect and outcomes**

The responsiveness analysis allows one to know which measure should be used to assess the outcome of any treatment in a specific population, and to calculate the sample size for any clinical trial (Gay et al., 2003; Katz et al., 1994; Liang, 1995). The results have demonstrated that the magnitude of treatment effect detected by the DASH was

higher than that observed by the physical examination measures for outcome assessment in carpal tunnel surgery. This concept has important implications in the interpretation of any outcome study in carpal tunnel surgery. The use of low-responsiveness measures can lead to an erroneous conclusion about the effectiveness of the treatment. In this paper, measures such as ROM and grip strength showed a negative treatment effect while other measures, such as the DASH and CTQ, showed improvement after treatment.

### **Construct validity and responsiveness analysis**

Validity is an expression of the degree to which an instrument measures what it purports to measure. Common methods to obtain construct validity include an examination of the logical relations that should exist between the measure under study with other measures and/or patterns of scores across groups of individuals (Medical Outcomes Trust Scientific Advisory Committee, 1997). The concept of detecting change over time is part of the validity assessment and therefore may be referred to as longitudinal construct validity (Gummenson et al., 2003). In fact, Gay et al. (2003) demonstrated in a longitudinal cohort study that the DASH presented a level of responsiveness higher than the SF-36 bp and SF-36 pf and a lower level than a specific disease instrument as CTQ. The same hypothesis was established in this paper to demonstrate the construct validity of the Spanish version of the DASH. The results were coincidental with this hypothesis as proof of construct validity of the Spanish version of the DASH. To sum up, the results of this paper present evidence of a good responsiveness level of the Spanish version of the DASH for outcome assessment in open carpal tunnel release, similar to

other equivalent versions, higher than the physical exam measures and with proof of good construct validity.

## **Acknowledgements**

The authors thank Prof Manuel Nogales for his statistical advice and Mrs Estefania Garcia Mesa, Prof of English language, for her contribution to this paper.

---

## References

1. Alonso J, Prieto L, Antó JM. La versión española del SF-36 Health Survey (cuestionario de Salud SF-36): un instrumento para la medida de los resultados clínicos. *Med Clin Barc.* 1995, 104:771–6.
2. Amadio P, Silverstein M, Ilstrup D, Schleck CD, Jensen LM. Outcome assessment for carpal tunnel surgery: the relative responsiveness of generic, arthritis-specific, disease specific, and physical examination measures. *J Hand Surg Am.* 1996, 21: 338–46.
3. Atroshi I, Gummenson C, Johnson R, Sprinchorn A. Symptoms, Disability, and Quality of Life in patients with carpal tunnel syndrome. *J Hand Surg Am.* 1999, 24: 398–404.
4. Beaton DE, Katz JN, Fossel AH, Wright JG, Tarasuk V, Bombardier C. Measuring the whole or the parts? Validity, reliability, and responsiveness of the Disabilities of the Arm, Shoulder and Hand outcome measure in different regions of the upper extremity. *J Hand Ther.* 2001, 14: 128–46.
5. De Smet L, De Kesel I, Degreef I, Debeer P. Responsiveness of the Dutch Version of the DASH as outcome measure for Carpal Tunnel Syndrome. *J Hand Surg Eur.* 2007, 32: 74–6.
6. Gay RE, Amadio PC, Johnson JC. Comparative Responsiveness of the Disabilities of the Arm, Shoulder, and Hand, the Carpal Tunnel Questionnaire, and the SF-36 to Clinical Change after Carpal Tunnel Release. *J Hand Surg Am.* 2003, 28: 250–4.

7. Greenslade JR, Mehta RL, Belward P, Warwick DJ. DASH and Boston questionnaire assessment of Carpal Tunnel Syndrome outcome: what is the responsiveness of an outcome questionnaire? *J Hand Surg Eur.* 2004, 29: 159–64.
8. Gummenson C, Atroshi I, Ekdahl C. The disabilities of arm, shoulder and hand (DASH) outcome questionnaire: longitudinal construct validity and measuring self-rated health change after surgery. *BMC Musculoskelet Disord.* 2003, 4: 11.
9. Hobby J, Watts C, Elliot D. Validity and responsiveness of the Patient Evaluation Measure as an outcome measure for Carpal Tunnel Syndrome. *J Hand Surg Eur.* 2005, 30: 350–4.
10. Katz JN, Gelberman RH, Wright EA, Lew RA, Liang MH. Responsiveness of self-reported and objective measures of disease severity in Carpal Tunnel Syndrome. *Med Care.* 1994, 32: 1127–33.
11. Kotsis S, Chung K. Responsiveness of the Michigan Hand Outcomes Questionnaire and the Disabilities of the Arm, Shoulder and Hand Questionnaire in Carpal Tunnel Surgery. *J Hand Surg Am.* 2005,30: 81–6.
12. Lee E, Chung M, Lo S. Construct validity of the Chinese version of the Disability of the Arm, Shoulder and Hand questionnaire (DASH-HKPWH). *J Hand Surg Eur.* 2005, 30: 29–34.
13. Levine D, Simmons BP, Mark JK et al., A self-administered questionnaire for the assessment of symptom severity and functional status in carpal tunnel syndrome. *J Bone Joint Surg Am.* 1993, 75: 1585–92.

- 
14. Liang MH. Evaluating measurement responsiveness. *J Rheumatol.*1995, 22: 1191–2. Medical Outcomes Trust Scientific Advisory Committee. Instrument review criteria. Medical Outcomes Trust Source Pages.1997, 6–9.
  15. Orfale A, Araujo P, Ferraz M, Natour J. Translation into Brazilian Portuguese, cultural adaptation and evaluation of the reliability of the Disabilities of the Arm, Shoulder and Hand questionnaire. *Braz J Med Biol Res.* 2005, 38: 293–302.
  16. Padua R, Padua E, Ceccarelli E, Romanini E et al., Italian version of the Disability of the Arm, Shoulder and Hand (DASH) questionnaire. Cross-cultural adaptation and validation. *J Hand Surg Eur.* 2003, 28: 179–86.
  17. Rosales RS, Bensenny-Delagdo E, Díez de La Lastra-Bosch I. Evaluation of the Spanish Version of the DASH and Carpal Tunnel Syndrome health-related quality of life instruments: cross-cultural adaptation process and reliability. *J Hand Surg Am.* 2002,27: 334–44.
  18. Themistocleus G, Goudelis G, Kyrou I, Chloros G et al., Translation into Greek, crosscultural adaptation and validation of the Disabilities of the Arm, Shoulder, and Hand Questionnaire (DASH). *J Hand Ther.* 2006, 19: 350–7.



## **Anexo D: artículo original**

### **“Reliability and Construct Validity of the Spanish Version of the 6-item CTS symptoms scale for outcomes assessment in Carpal Tunnel Syndrome”**

R.S.Rosales<sup>1</sup>, Y. Martín Hidalgo<sup>2</sup>, L. Reboso Morales<sup>2</sup>, Isam Atroshi<sup>3</sup>

<sup>1</sup>Unit for hand & Microsurgery, GECOT, María del Cristo Ossuna 20, La Laguna, Tenerife, 38204, Tenerife.

<sup>2</sup>Orthopaedic Surgery Service, University Hospital of "La Candelaria", Carretera del Rosario s/n, La Laguna, Tenerife 38110, Spain

<sup>3</sup>Departments of Clinical Sciences Lund University, Lund, and Orthopedics Hässleholm-Kristianstad, Hässleholm Hospital, Hässleholm. Sweden.

Artículo aceptado en fase de revisión.

## **BACKGROUND**

The purpose of this paper was to assess the reliability and construct validity of the Spanish version of the 6-item carpal tunnel syndrome (CTS) symptoms scale (CTS-6).

## **METHODS**

In this cross-sectional study 40 patients with CTS, based on clinical and neurophysiological criteria, completed the standard Spanish versions of the CTS-6, QuickDASH and EQ-5D on two occasions with a 1-week interval. Internal-consistency reliability was assessed with the Cronbach alpha coefficient and test-retest reliability with the intraclass correlation coefficient (ICC). For assessing construct validity it was hypothesized that the CTS-6 would have a strong positive correlation with the QuickDASH and a moderate negative correlation with the EQ-5D Index, analyzed with the Pearson correlation coefficient ( $r$ ).

## **RESULTS**

The standard Spanish version of the CTS-6 presented good internal consistency (alpha 0.81) and test-retest reliability (ICC 0.85). The correlations were concordant with the a priori formulated construct hypothesis ( $r$  0.69 and -0.45, respectively).

## **CONCLUSIONS**

The standard Spanish version of the 6-item CTS symptoms scale has a high level of internal consistency and test-retest reliability and good construct validity for outcomes assessment in CTS. The CTS-6 will be useful to clinicians and researchers in Spanish

speaking parts of the world. The use of standardized outcome measures across countries also will facilitate comparison of research results in carpal tunnel syndrome.

## **BACKGROUND**

The use of disease-specific measures of patient-reported outcomes (PRO) has grown in clinical research. Carpal tunnel syndrome (CTS) is one of the most frequent conditions managed at hand surgery services. The CTS questionnaire developed by Levine et al. [1] has been among the most widely used PRO measures during the last two decades. The CTS questionnaire consists of two scales: symptoms severity (SS) (11 items) and functional status (FS). Using Factor analysis and Items Response Theory methodology, Atroshi et al. [2] developed a short version of the CTS SS-scale with only 6 items with the purpose of reducing respondent burden while maintaining the psychometric properties. They have demonstrated that the new brief version possessed a good level of reliability, validity and responsiveness [2-4]. Because a Spanish version of the 11-item symptom severity scale was already available [5], a new Spanish version of the shorter version (CTS-6) was released [6]. No new evidence has been reported about the reliability and validity of the CTS-6 since the first description done by Atroshi et al.[2] in 2009.

The purpose of this study was to assess the reliability and construct validity of the Spanish version of the 6-items CTS symptoms scale for outcomes assessment in CTS.

## **METHODS**

### **Study Population.**

The institutional review board of our medical centre approved the protocol. Informed consent was obtained from each subject before participation. The study population consisted of 40 consecutive patients (26 women), mean age 54 (SD 11) years, from the waiting list for carpal tunnel release surgery in the National Health System in Spain. The inclusion criteria were: 1) numbness or tingling with or without pain in at least 2 of the 4 radial digits [7,8], (2) increased symptoms with carpal tunnel provocation tests (positive Phalen and/or reverse Phalen test) [8], (3) symptoms for over two months [7], (4) failure of conservative treatment [7], and nerve conduction test showing median neuropathy at the wrist (distal motor latency > 4.5 milliseconds, wrist-digit sensory latency > 3.5 milliseconds, or sensory conduction velocity at the carpal tunnel segment < 40 meters/second) [9,10] . The exclusion criteria were: clinical or electrophysiological signs of proximal nerve compression, diabetes or any metabolic disease, rheumatoid arthritis or other general inflammatory diseases [7,8,11] .

**Clinical Design.** A cross-sectional study.

**Outcomes measures.** The standard Spanish versions of the CTS-6 [6], the 11-item disabilities of the arm, shoulder and hand (QuickDASH) questionnaire ([www.iwh.on.ca](http://www.iwh.on.ca)) and the Euroqol 5D Index (EQ-5D) [12] were completed by the patients at the outpatient clinic.

The CTS-6 measured symptoms severity related to CTS. It is constituted by 6 items. Five of the 6 items in the CTS-6 have similar item text as the corresponding items in the

11-item symptom severity scale (original CTS-SS questionnaire) and the remaining item (the result of merger of 2 symptom severity scale items) has text from the 2 items. The CTS-6 has, however, a completely different and improved layout. The scoring is similar to that for the 11-item symptom severity score; for each patient the item responses are scored from 1 (best) to 5 (worst) and then averaged for the 6 items to yield a CTS-6 score (only 1 missing item response is allowed).

The QuickDASH is the shorter version of the DASH PRO measure developed for measuring “upper extremities disability”. It is constituted by 11 items and it is scored from 0 (best =lowest disability) to 100 (worst = highest disability). At least 10 of the 11 items must be completed for a score to be calculated. Each item is scored 1 to 5 and the assigned values for all completed items are summed and averaged, producing a score of 1 to 5. This value is then transformed to a score of 0 to 100 by subtracting one and multiplying by 25. This transformation is done to make the score easier to compare to other measures scaled on a 0-100 scale.

The EQ-5D generic health index comprises a five-part questionnaire and a visual analogue self-rating scale. The five dimensions in the EQ-5D are mobility (MO), self-care (SC), usual activities (UA), pain/discomfort (PD), and anxiety/depression (AD). The possible range for each of the dimension variables is 1 to 3, where 1= no problems, 2 = moderate problems, and 3 = extreme problems. Once the analytical dataset has appropriately named dimension variables, the EQ-5D Index was generated using the EQ- 5DTM Scoring algorithm for Excel

(<http://www.ahrq.gov/professionals/cliniciansproviders/resources/rice/EQ5Dscore.html>).

The EQ-5D Index scores range from -0.59 (worst) to 1 (best health). The index will not be calculated when responses are missing for one or more of the dimensions. No missing items from the three PRO instruments were observed in this study. For assessing construct validity it was hypothesized that the CTS-6 scores would have a moderate to strong positive correlation with the QuickDASH scores and a low to moderate negative correlation with the EQ-5D Index. For assessing test-retest reliability a second self administration of the CTS-6 was done at the clinic 1 week after the first administration.

**Data analysis.** Internal-consistency reliability was assessed with the Cronbach alpha coefficient ( $\alpha > 0.7$  indicates good internal consistency). Test-retest reliability was assessed with the intraclass correlation coefficient (ICC) and by comparing the mean CTS-6 scores for the two consecutive administrations with the t-test. For the test-retest reliability mean difference analysis, a sample of 19 individuals will be needed to detect an important clinical difference of 0.9 in the CTS-6 scores, assuming a SD of 0.7 [2] , two-sided test, power of 80%, and a level of significance of 0.05. The construct validity hypotheses were analyzed with the Pearson correlation coefficient ( $r$ ), using a level 0.01 for statistical significance.

## RESULTS

Reliability: the Cronbach alpha coefficient was 0.81 and the ICC was 0.85 ( $p < 0.001$ ). The mean difference in the CTS-6 scores for the two administrations was 0.09 (95% confidence interval -0.07 to 0.26,  $p = 0.269$ ) (Table 1). Construct Validity: there was a

strong positive correlation between the scores for the CTS-6 scale and the QuickDASH ( $r = 0.685$ ,  $p < 0.001$ ), and a moderate negative correlation with the EQ-5D Index ( $r = -0.45$ ,  $p = 0.003$ ).

## **DISCUSSION**

The results have demonstrated that the CTS-6 PRO measure presented a good level of internal consistency and high level of test-retest reliability with a mean difference of the scores in two administrations, 1 week apart, which was not statistically different from zero and with an excellent level of intraclass correlation. Besides, the correlations were concordant with the construct hypothesis formulated a priori as a proof of good construct validity.

One of the measurement property of the PRO instruments included by the Medical Outcomes Trust in the instruments review criteria have been the “respondent burden” defines as the time, energy, and other demands placed on those to whom the instrument is administered [13] The CTS-6 was developed to improve patient acceptance, to increase the level of responses and consequently the efficacy of the PRO instrument while the shorter scale maintains the psychometric properties [3].

In this paper, the Spanish CTS-6 presented a Cronbach alpha coefficient of 0.81 and an ICC of 0.85. Similar results have been reported with the original CTS-6 (Cronbach alpha= 0.86, ICC= 0.95) [2]. The mean difference of scores in two administration in the original CTS-6 was 0.03 (95% CI -0.07 to 0.12) [2]. In our sample the mean difference scores was 0.09 (95% CI -0.07 to 0.26). Comparing to the reliability of the longer Spanish version (11-items SS scale), with the same “washout time” of 1 week,

the results were similar too (mean differences scores = 0.18, 95% CI -0.16 to 0.53,  $p=0.248$ ) [5] . Consequently, the Spanish CTS-6 presents a high level of reliability similar to the original CTS-6 and the Spanish CTS-11.

There are different ways to analyze validity of a health outcome measure: content validity, criterion-validity, and construct validity [13-15]. Common methods to obtain construct-related validity include an examination of the logical relationship that should exist between that measure and other measures and/or patterns of scores across groups of individuals. Testing for construct validity involves assessing both theory and method simultaneously. Therefore, it should include the hypothesis that can demonstrate the proposed construct (theory) [14,15]. Many factors should be considered when choosing hypotheses. The most important factor is the specific dimension of health or concept that is intended to be measured by using a patient-completed questionnaire and the way or direction of scoring of every instrument studied. Atroshi et al. [2] demonstrated a convergent validity by a strong correlation between the original CTS-6 scores and the QuickDASH scores ( $r=0.7$ ). In this study we have observed a very similar correlation coefficient ( $r=0.69$ ) between the Spanish CTS-6 and QuickDASH. Besides, the expected negative low to moderate correlation between CTS-6 and the EQ 5D-Index ( $-0.46$ ) constitutes another proof of the validity of the CTS-6 (severe CTS symptoms should correlate with lower quality of Life). To our knowledge no previous study has used the EQ 5D-Index in construct validity analysis. The results of this study showed that the prespecified construct hypothesis was established.

Based on the COSMIN (Consensus-based Standards for the Selection of health status Measurement Instruments) check list [14,15] for assessing the quality of studies on

measurements properties of PRO Instruments, the construct validity analysis of this paper met 10 from 10 design requirements as standards for good methodological quality: description of how the missing items were handled, percentage of missing values, sample size calculation, the hypothesis regarding correlations was formulated a priori before the data collections, the expected directions of the correlations were included in the hypothesis, the expected absolute or relative magnitude of correlations were include in the hypothesis, there was an adequate description of comparator instruments and their properties, and finally adequate clinical design and statistical method for the hypothesis to be tested.

Limitations of this study is the lack of responsiveness analysis. Further studies regarding responsiveness and minimal clinically important difference (MCID) are recommended to complete the analysis of the measurements properties of the Spanish CTS-6.

## **CONCLUSSIONS**

This study have demonstrated that the Spanish CTS-6 PRO measure has good reliability and construct validity for outcomes assessment in CTS. The CTS-6 will be useful to clinicians and researchers in Spanish speaking parts of the world. The use of standardized outcome measures across countries also will facilitate comparison of research results in carpal tunnel syndrome.

## **COMPETING INTERESTS**

There are not financial or non- financial competing interests to declare in relation to this manuscript.

## **AUTHORS CONTRIBUTIONS**

RSR & IA were involved as directors of the study. They made the conception and design of the study, as well as, they were involved in the analysis and the interpretation of the data.

LRB & YMH were the researcher who did the clinical exam, they enrolled the patients based on the inclusion and exclusion criteria and they collected the data. All the authors read and approved the final version of the manuscript.

---

## REFERENCES

1. Levine D, Simmons B P, Mark J K, et al. . A self-administered questionnaire for the assessment of symptom severity and functional status in carpal tunnel syndrome. *J Bone J Surg Am.* 1993; 75: 1585-92.
2. Atroshi I, Lyren PE, Gummesson C. The 6-item CTS symptoms scale: a brief outcomes measure for carpal tunnel syndrome. *Qual Life Res.* 2009; 18: 347-58.
3. Atroshi I, Lyren PE, Ornstein E, Gummesson C. The six-item CTS symptoms scale and palmar pain scale in carpal tunnel syndrome. *J Hand Surg Am.* 2011; 36: 788-94.
4. Lyren PE, Atroshi I. Using item response theory improved responsiveness of patient reported outcomes measures in carpal tunnel syndrome. *J Clin Epidemiol.* 2012; 65:325- 34.
5. Rosales RS, Delgado EB, Diez de la Lastra-Bosch I. Evaluation of the Spanish version of the DASH and carpal tunnel syndrome health-related quality-of-life instruments: cross- cultural adaptation process and reliability. *J Hand Surg Am.* 2002; 27: 334-43.
6. Rosales RS, Atroshi I. Spanish versions of the 6-item carpal tunnel syndrome symptoms scale (CTS-6) and palmar pain scale. *J Hand Surg Eur.* 2013; 38:550-1.
7. Atroshi I, Gummenson C, Johnson R, Sprinchorn A . Symptoms, Disability, and Quality of Life in patients with carpal tunnel syndrome. *J Hand Surg Am.* 1999; 24 : 398-404.

8. Gay R E, Amadio P C, Johnson J C . Comparative Responsiveness of the Disabilities of the Arm, Shoulder, and Hand, the Carpal Tunnel Questionnaire, and the SF-36 to Clinical Change after Carpal Tunnel Release. *J Hand Surg Am.* 2003; 28 : 250-54.
9. Atroshi I, Larsson GU, Ornstein E, Hofer M, Johnsson R, Ranstam J. Outcomes of endoscopic surgery compared with open surgery for carpal tunnel syndrome among employed patients: randomised controlled trial. *BMJ.* 2006; 332:1473
10. Kimura J. *Electrodiagnosis in diseases of nerve and muscle: principles and practice.* 2nd ed. Philadelphia: FA Davis; 1989.
11. Amadio P, Silverstein M, Ilstrup D, Schleck CD, Jensen LM . Outcome assessment for carpal tunnel surgery: the relative responsiveness of generic, arthritis-specific, disease- specific, and physical examination measures. *J Hand Surg Am.* 1996; 21: 338-46.
12. Badia X, Roset M, Montserrat S, Herdman M, Segura A. The Spanish version of EuroQol: a description and its applications. *European Quality of Life scale. Med Clin (Barc).* 1999; 112 Suppl 1:79-85.
13. Medical Outcomes Trust Scientific Advisory Committee . Instrument review criteria. *Medical Outcomes Trust Source Pages.* 1997: 6-9
14. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, Bouter LM, de Vet HC. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol.* 2010;10:22. doi: 10.1186/1471-2288-10-22.

15. 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: 651-57.



## **Anexo E: CUESTIONARIO SF-36**

### **INSTRUCCIONES:**

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.

Conteste cada pregunta tal como se indica. Si no está seguro/a de cómo responder a una pregunta, por favor conteste lo que le parezca más cierto.

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

- 1 Excelente
- 2 Muy Buena
- 3 Buena
- 4 Regular
- 5 Mala

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

- 1 Mucho mejor ahora que hace un año
- 2 Algo mejor ahora que hace un año
- 3 Más o menos igual que hace un año
- 4 Algo peor ahora que hace un año
- 5 Mucho peor ahora que hace un año

**LAS SIGUIENTES PREGUNTAS SE REFIEREN A ACTIVIDADES  
O COSAS QUE USTED PODRÍA HACER EN UN DÍA NORMAL.**

3. Su salud actual, ¿le limita para hacer **esfuerzos intensos**, tales como correr, levantar objetos pesados, o participar en deportes agotadores?

- 1 " Sí, me limita mucho
- 2 " Sí, me limita un poco
- 3 " No, no me limita nada

4. Su salud actual, ¿le limita para hacer **esfuerzos moderados**, como mover una mesa, pasar la aspiradora, jugar a los bolos o caminar más de una hora?

- 1 " Sí, me limita mucho
- 2 " Sí, me limita un poco
- 3 " No, no me limita nada

5. Su salud actual, ¿le limita para **coger o llevar la bolsa de la compra**?

- 1 Sí, me limita mucho
- 2 Sí, me limita un poco
- 3 No, no me limita nada

6. Su salud actual, ¿le limita para **subir varios pisos** por la escalera?

- 1 Sí, me limita mucho
- 2 Sí, me limita un poco
- 3 No, no me limita nada

7. Su salud actual, ¿le limita para **subir un solo piso** por la escalera?

- 1 Sí, me limita mucho
- 2 Sí, me limita un poco
- 3 No, no me limita nada

8. Su salud actual, ¿le limita para **agacharse o arrodillarse**?

- 1 Sí, me limita mucho
- 2 Sí, me limita un poco
- 3 No, no me limita nada

9. Su salud actual, ¿le limita para caminar **un kilómetro o más**?

- 1 Sí, me limita mucho
- 2 Sí, me limita un poco
- 3 No, no me limita nada

10. Su salud actual, ¿le limita para caminar **varias manzanas** (varios centenares de metros)?

- 1 " Sí, me limita mucho
- 2 " Sí, me limita un poco
- 3 " No, no me limita nada

11. Su salud actual, ¿le limita para caminar **una sola manzana** (unos 100 metros)?

- 1 " Sí, me limita mucho
- 2 " Sí, me limita un poco
- 3 " No, no me limita nada

12. Su salud actual, ¿le limita para **bañarse o vestirse por sí mismo?**

- 1 " Sí, me limita mucho
- 2 " Sí, me limita un poco
- 3 " No, no me limita nada

**LAS SIGUIENTES PREGUNTAS SE REFIEREN A PROBLEMAS  
EN SU TRABAJO O EN SUS ACTIVIDADES COTIDIANAS.**

13. Durante las 4 últimas semanas, ¿tuvo que **reducir el tiempo** dedicado al trabajo o a sus actividades cotidianas, a causa de su salud física?

1 " Sí

2 " No

14. Durante las 4 últimas semanas, ¿hizo **menos** de lo que hubiera querido hacer, a causa de su salud física?

1 " Sí

2 " No

15. Durante las 4 últimas semanas, ¿tuvo que **dejar de hacer algunas tareas** en su trabajo o en sus actividades cotidianas, a causa de su salud física ?

1 " Sí

2 " No

16. Durante las 4 últimas semanas, ¿tuvo **dificultad** para hacer su trabajo o sus actividades cotidianas (por ejemplo, le costó más de lo normal), a causa de su salud física?

- 1 " Sí
- 2 " No

17. Durante las 4 últimas semanas, ¿tuvo que **reducir el tiempo** dedicado al trabajo o a sus actividades cotidianas, a causa de algún problema emocional (como estar triste, deprimido, o nervioso)?

- 1 " Sí
- 2 " No

18. Durante las 4 últimas semanas, ¿ **hizo menos** de lo que hubiera querido hacer, a causa de algún problema emocional (como estar triste, deprimido, o nervioso)?

- 1 " Sí
- 2 " No

19. Durante las 4 últimas semanas, ¿no hizo su trabajo o sus actividades cotidianas tan **cuidadosamente** como de costumbre, a causa de algún problema emocional (como estar triste, deprimido, o nervioso)?

- 1 " Sí
- 2 " No

20. 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?

- 1 " Nada
- 2 " Un poco
- 3 " Regular
- 4 " Bastante
- 5 " Mucho

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

- 1 " No, ninguno
- 2 " Sí, muy poco
- 3 " Sí, un poco
- 4 " Sí, moderado
- 5 " Sí, mucho
- 6 " Sí, muchísimo

22. 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)?

- 1 " Nada
- 2 " Un poco
- 3 " Regular
- 4 " Bastante
- 5 " Mucho

**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.**

23. Durante las 4 últimas semanas, ¿cuánto tiempo se sintió lleno de vitalidad?

- 1 " Siempre
- 2 " Casi siempre
- 3 " Muchas veces
- 4 " Algunas veces
- 5 " Sólo alguna vez
- 6 " Nunca

24. Durante las 4 últimas semanas, ¿cuánto tiempo estuvo muy nervioso?

- 1 " Siempre
- 2 " Casi siempre
- 3 " Muchas veces
- 4 " Algunas veces
- 5 " Sólo alguna vez
- 6 " Nunca

25. Durante las 4 últimas semanas, ¿cuánto tiempo se sintió tan bajo de moral que nada podía animarle?

- 1 " Siempre
- 2 " Casi siempre
- 3 " Muchas veces
- 4 " Algunas veces
- 5 " Sólo alguna vez
- 6 " Nunca

26. Durante las 4 últimas semanas ¿cuánto tiempo se sintió calmado y tranquilo?

- 1 " Siempre
- 2 " Casi siempre
- 3 " Muchas veces
- 4 " Algunas veces
- 5 " Sólo alguna vez
- 6 " Nunca

27. Durante las 4 últimas semanas, ¿cuánto tiempo tuvo mucha energía?

- 1 " Siempre
- 2 " Casi siempre
- 3 " Muchas veces
- 4 " Algunas veces
- 5 " Sólo alguna vez
- 6 " Nunca

28. Durante las 4 últimas semanas, ¿cuánto tiempo se sintió desanimado y triste?

- 1 " Siempre
- 2 " Casi siempre
- 3 " Muchas veces
- 4 " Algunas veces
- 5 " Sólo alguna vez
- 6 " Nunca

29. Durante las 4 últimas semanas, ¿cuánto tiempo se sintió agotado?

- 1 " Siempre
- 2 " Casi siempre
- 3 " Muchas veces
- 4 " Algunas veces
- 5 " Sólo alguna vez
- 6 " Nunca

30. Durante las 4 últimas semanas, ¿cuánto tiempo se sintió feliz?

- 1 " Siempre
- 2 " Casi siempre
- 3 " Muchas veces
- 4 " Algunas veces
- 5 " Sólo alguna vez
- 6 " Nunca

31. Durante las 4 últimas semanas, ¿cuánto tiempo se sintió cansado?

- 1 " Siempre
- 2 " Casi siempre
- 3 " Muchas veces
- 4 " Algunas veces
- 5 " Sólo alguna vez
- 6 " Nunca

32. 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)?

- 1 " Siempre
- 2 " Casi siempre
- 3 " Algunas veces
- 4 " Sólo alguna vez
- 5 " Nunca

**POR FAVOR, DIGA SI LE PARECE CIERTA O FALSA CADA  
UNA DE LAS SIGUIENTES FRASES.**

33. Creo que me pongo enfermo más fácilmente que otras personas.

- 1 " Totalmente cierta
- 2 " Bastante cierta
- 3 " No lo sé
- 4 " Bastante falsa
- 5 " Totalmente falsa

34. Estoy tan sano como cualquiera.

- 1 " Totalmente cierta
- 2 " Bastante cierta
- 3 " No lo sé
- 4 " Bastante falsa
- 5 " Totalmente falsa

35. Creo que mi salud va a empeorar.

- 1 " Totalmente cierta
- 2 " Bastante cierta
- 3 " No lo sé
- 4 " Bastante falsa
- 5 " Totalmente falsa

36. Mi salud es excelente.

1 " Totalmente cierta

2 " Bastante cierta

3 " No lo sé

4 " Totalmente cierta

5 " Totalmente falsa



## Anexo F: Versión Española del Instrumento DASH

Por favor puntúe su habilidad o capacidad para realizar las siguientes actividades durante la última semana. Para ello marque con un círculo el número apropiado para cada respuesta.

	No Dificultad	Dificultad Leve	Dificultad Moderada	Mucha Dificultad	Imposible de Realizar
1. Abrir un bote de cristal nuevo	1	2	3	4	5
2. Escribir	1	2	3	4	5
3. Preparar la comida	1	2	3	4	5
4. Girar la llave para abrir la puerta o encender el coche	1	2	3	4	5
5. Empujar y abrir una puerta pesada	1	2	3	4	5
6. Colocar un objeto en estanterías situadas por encima de su cabeza	1	2	3	4	5
7. Realizar tareas duras de la casa como fregar el piso, limpiar paredes y cristales etc	1	2	3	4	5
8. Bañarse o vestirse	1	2	3	4	5
9. Hacer las camas	1	2	3	4	5
10. Cargar una bolsa del supermercado o maletín	1	2	3	4	5
11. cargar un objeto	1	2	3	4	5

<b>pesado ( más de 5Kg)</b>					
<b>12. Cambiar una bombilla del techo</b>	1	2	3	4	5
<b>13. Lavarse y secarse el pelo</b>	1	2	3	4	5
<b>14. Lavarse la espalda</b>	1	2	3	4	5
<b>15. Ponerse un jersey o un suéter</b>	1	2	3	4	5
<b>16. Usar un cuchillo para cortar la comida</b>	1	2	3	4	5
<b>17. Actividades de entretenimiento que requieren poco esfuerzo como hacer punto, coser, jugar a las cartas o al dominó</b>	1	2	3	4	5
<b>18. Actividades que requieren algo de esfuerzo para su brazo, hombro o mano, como usar un martillo, jugar al golf, al tenis o a la petanca</b>	1	2	3	4	5
<b>19. Actividades que requieren que su brazo se mueva libremente como nadar</b>	1	2	3	4	5
<b>20. Conducir</b>	1	2	3	4	5
<b>21. Actividad sexual</b>	1	2	3	4	5
<b>22. . ¿Durante la última semana, sus problemas del hombro, brazo o mano han interferido con su actividad social habitual con la</b>	1	2	3	4	5

familia, sus amigos o  
compañeros de trabajo?

<b>23. ¿Durante la última semana ha tenido usted dificultad para realizar su trabajo u otras actividades diarias debido a problemas en su hombro, brazo o mano?</b>	1	2	3	4	5
---	---	---	---	---	---

Por favor ponga puntuación a la gravedad o severidad de los siguientes síntomas

	Ninguno	Leve	Moderado	Grave	Muy Grave
<b>24. Dolor en el hombro, brazo o mano</b>	1	2	3	4	5
<b>25. Dolor en el hombro, brazo o mano cuando realiza cualquier actividad específica</b>	1	2	3	4	5
<b>26. Sensación de calambres, hormigueos o de electricidad en el hombro, brazo o mano</b>	1	2	3	4	5
<b>27. Debilidad o falta de fuerza en el hombro, brazo y mano</b>	1	2	3	4	5
<b>28. Rigidez o falta de movilidad en el hombro, brazo y mano</b>	1	2	3	4	5
					<b>Dificultad</b>

	No	Leve	Moderada	Grave	extrema que me impide dormir
<b>29. ¿Durante la última semana, ha tenido dificultad para dormir debido al dolor en el hombro, brazo o mano?</b>	1	2	3	4	5
	Completamente falso	Falso	No lo sé	Cierto	Completamente cierto
<b>30. Me siento con menos capacidad, confianza y útil debido a mi problema con el hombro, brazo o mano</b>	1	2	3	4	5

### Módulo de Trabajo (Opcional)

Las siguientes preguntas se refieren al impacto que tiene su problema del brazo, hombro o mano en su capacidad para trabajar (incluyendo las tareas de la casa si ese es su trabajo principal).

Por favor, indique cuál es su trabajo/ocupación: -----

Yo no trabajo (usted puede pasar por alto esta sección).

Marque con un círculo el número que describa mejor su capacidad física en la semana pasada. ¿Tuvo usted alguna dificultad .....

	Ninguna dificultad	Dificultad leve	Dificultad moderada	Mucha dificultad	Imposible
1. Para usar su técnica habitual de trabajo?	1	2	3	4	5
2. Para hacer su trabajo habitual debido al dolor del hombro, brazo o mano?	1	2	3	4	5
3. Para realizar su trabajo tan bien como le gustaría?	1	2	3	4	5
4. para emplear la cantidad habitual de tiempo en su trabajo?	1	2	3	4	5

### Actividades especiales deportes/músicos (Opcional)

Las preguntas siguientes hacen referencia al impacto que tiene su problema en el brazo, hombro o mano para tocar su instrumento musical, practicar su deporte o ambos. Si usted practica más de un deporte o toca más de un instrumento (o hace ambas cosas), por favor conteste con respecto a la actividad que sea más importante para usted. Por favor, indique el deporte o instrumento que sea más importante para usted.

¿Tuvo alguna dificultad?

	Ninguna dificultad	Dificultad leve	Dificultad moderada	Mucha dificultad	Imposible
1. Para usar su técnica habitual al tocar su instrumento o practicar su deporte?	1	2	3	4	5
2. Para tocar su instrumento habitual o practicar su deporte debido a dolor en el brazo, hombro, mano?	1	2	3	4	5
3. Para tocar su instrumento o practicar su deporte tan bien como le gustaría?	1	2	3	4	5
4. Para emplear la cantidad de tiempo habitual para tocar su instrumento o practicar su deporte?	1	2	3	4	5

## **Anexo G: QUICKDASH**

### **INSTRUCCIONES:**

Este cuestionario le pregunta sobre sus síntomas así como su capacidad para realizar ciertas actividades o tareas. Por favor conteste cada pregunta basándose en su condición o capacidad durante la última semana. Para ello marque un círculo en el número apropiado. Si usted no tuvo la oportunidad de realizar alguna de las actividades durante la última semana, por favor intente aproximarse a la respuesta que considere que sea la más exacta. No importa que mano o brazo usa para realizar la actividad; por favor conteste basándose en la habilidad o capacidad y como puede llevar a cabo dicha tarea o actividad. Por favor puntúe su habilidad o capacidad para realizar las siguientes actividades durante la última semana. Para ello marque con un círculo el número apropiado para cada respuesta.

	Ninguna Dificultad	Dificultad Leve	Dificultad Moderada	Dificultad Severa	Imposible de realizar
1. Abrir un bote de cristal nuevo	1	2	3	4	5
2. Realizar tareas duras de la casa (fregar el suelo, limpiar paredes etc)	1	2	3	4	5
3. Cargar una bolsa de supermercado o maletín	1	2	3	4	5
4. Lavarse la espalda	1	2	3	4	5
5. Usar un cuchillo para cortar comida	1	2	3	4	5
6. Actividades de entretenimiento que requieren algo de esfuerzo o impacto para su brazo, hombro, mano (golf, martillar, tenis, petanca)	1	2	3	4	5
	No, para nada	Un poco	Regular	Bastante	Mucho
7. ¿Durante la última semana su problema en el hombro, brazo o mano ha interferido con sus actividades sociales normales, con la familia, amigos, vecinos o grupos?	1	2	3	4	5
	No	Un poco	Moderadamente Limitado	Bastante	Imposible de realizar
8. ¿Durante la última semana ha tenido usted dificultad para realizar su trabajo u otras actividades cotidianas debido a su problema en el brazo, hombro o mano?	1	2	3	4	5
	Ninguno	Leve	Moderado	Grave	Muy Grave
9. Dolor en brazo, hombro, mano	1	2	3	4	5
10. Sensación de calambres, hormigueos o alfilerazos en su brazo, hombro o mano	1	2	3	4	5
	No	Media	Moderada	Grave	Dificultad extrema
11. Dificultad para dormir en la última semana debido al dolor en su brazo, hombro, mano?	1	2	3	4	5

Cálculo de la puntuación del “Quick Dash” (Discapacidad/Síntomas) =  $(\frac{\text{suma de } n \text{ respuestas}}{n} - 1) \times 25$ , donde n es igual al número respuestas completadas.

La puntuación del “Quick Dash” no puede ser calculada si hay más de 1 ítem sin contestar

## Anexo H: Cuestionario de Salud Euroqol-5D

### Sistema Descriptivo.

Marque con una cruz la respuesta de cada apartado que mejor describa su estado de salud en el día de HOY.

#### Movilidad

- No tengo problemas para caminar
- Tengo algunos problemas para caminar
- Tengo que estar en la cama


#### Cuidado personal

- No tengo problemas con el cuidado personal
- Tengo algunos problemas para lavarme o vestirme
- Soy incapaz de lavarme o vestirme


**Actividades cotidianas** (p.ej, trabajar, estudiar, hacer las tareas domésticas, actividades familiares o actividades durante el tiempo libre)

- No tengo problemas para realizar mis actividades cotidianas
- Tengo algunos problemas para realizar mis actividades cotidianas
- Soy incapaz de realizar mis actividades cotidianas


**Dolor/malestar**

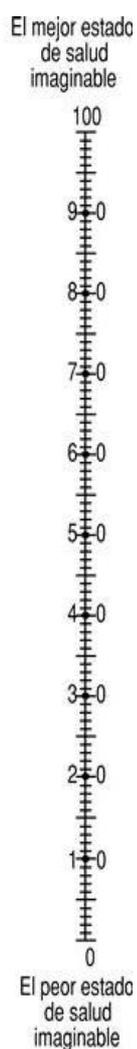
- No tengo dolor ni malestar
- Tengo moderado dolor o malestar Tengo mucho dolor o malestar


**Ansiedad/Depresión**

- No estoy ansioso ni deprimido
- Estoy moderadamente ansioso o deprimido
- Estoy muy ansioso o deprimido


### Escala Visual analógica del EQ-5D

Nos gustaría que nos indicara en esta escala, en su opinión, lo bueno o malo que es su estado de salud en el día de HOY. Por favor dibuje una línea desde el casillero donde dice "su estado de salud hoy" hasta el punto del termómetro que en su opinión indique lo bueno o malo que es su estado de salud en el día de HOY



#### Termómetro Euroqol de Autovaloración del Estado de Salud

Para ayudar a la gente a describir lo bueno o malo que es su estado de salud hemos dibujado una escala parecida a un termómetro en el cual se marca con un 100 el mejor estado de salud que pueda imaginarse y con un 0 el peor estado de salud que pueda imaginarse.

Nos gustaría que nos indicara en esta escala, en su opinión, lo bueno o malo que es su estado de salud en el día de HOY. Por favor dibuje una línea desde el casillero donde dice " su estado de salud hoy" hasta el punto del termómetro que en su opinión indique lo bueno o malo que es su estado de salud en el día de HOY.



# Anexo I: Versión Española del Instrumento CTS

## Escala del Estado Funcional

¿En un día normal durante las últimas dos semanas, ha tenido alguna dificultad para realizar las actividades enumeradas a continuación, debido a síntomas de la mano o muñeca? Por favor marque con un círculo en cada línea el número que mejor defina su habilidad o capacidad para realizar la actividad.

	No Dificultad	Dificultad Leve	Dificultad Moderada	Mucha Dificultad	Imposible de Realizar
A. Escribir	1	2	3	4	5
B. Abrocharse los botones de la ropa	1	2	3	4	5
C. Sujetar un libro mientras lee	1	2	3	4	5
D. Sujetar el teléfono	1	2	3	4	5
E. Abrir botes	1	2	3	4	5
F. Hacer las tareas de la casa	1	2	3	4	5
G. Llevar una bolsa de la compra	1	2	3	4	5
H. Bañarse o vestirse	1	2	3	4	5

## Escala de Severidad de Síntomas

Las siguientes preguntas se refieren a sus síntomas durante un periodo normal de 24 h durante las últimas 2 semanas.

- ¿Cuál es la intensidad del dolor en la mano o muñeca por la noche?
  1. No tengo dolor en la mano o muñeca por la noche
  2. dolor leve
  3. dolor moderado
  4. dolor severo
  5. dolor muy severo
  
- ¿Con qué frecuencia le despertó el dolor en la mano o muñeca durante una noche normal en las últimas 2 semanas?
  1. Nunca
  2. Una vez
  3. Dos o tres veces
  4. Cuatro o cinco veces
  5. Más de cinco veces
  
- ¿Tiene usted normalmente dolor en su mano o muñeca durante el día?
  1. Nunca tengo dolor durante el día
  2. Tengo dolor leve durante el día
  3. Tengo dolor moderado durante el día
  4. Tengo dolor severo durante el día
  5. Tengo dolor muy severo durante el día

- 
- ¿Con qué frecuencia tiene usted dolor en su mano o muñeca durante el día?
    1. Nunca
    2. Una o dos veces
    3. Tres a cinco veces
    4. Más de cinco veces
    5. El dolor es constante
  
  - ¿Cuánto tiempo le dura, como promedio, un episodio de dolor durante el día?
    1. Nunca tengo dolor durante el día
    2. 10 minutos o menos
    3. De 10 a 60 minutos
    4. Más de 60 minutos
    5. El dolor es constante
  
  - ¿Tiene usted adormecimiento (pérdida de sensación) en su mano?
    1. No
    2. Tengo un leve adormecimiento
    3. Tengo un adormecimiento moderado
    4. Tengo un adormecimiento grave
    5. Tengo un adormecimiento muy grave
  
  - ¿Tiene usted debilidad o falta de fuerza en su mano o muñeca?
    1. no
    2. debilidad leve
    3. debilidad moderada
    4. debilidad grave
    5. debilidad muy grave

- ¿Tiene usted sensación de hormigueos en su mano?
  1. No
  2. Hormigueos de carácter leve
  3. Hormigueos de carácter moderado
  4. Hormigueos de carácter grave
  5. Hormigueos de carácter muy grave
  
- ¿Cómo de grave es el adormecimiento (pérdida de sensación) o los hormigueos, por la noche?
  1. No tengo ni hormigueos ni adormecimiento por la noche
  2. Leve
  3. Moderado
  4. Grave
  5. Muy grave
  
- ¿Con qué frecuencia le despertó el adormecimiento u hormigueo en la mano durante una noche normal en las últimas 2 semanas?
  1. Nunca
  2. Una vez
  3. Dos o tres veces
  4. Cuatro a cinco veces
  5. Más de cinco veces
  
- ¿Tiene usted dificultad para agarrar y usar objetos pequeños como llaves, bolígrafo, etc?
  1. No
  2. Dificultad leve
  3. Dificultad moderada
  4. Dificultad grave
  5. Dificultad muy grave

## Anexo J: Versión Española del Instrumento 6-CTS

Las siguientes preguntas se refieren a sus síntomas durante un periodo normal de 24 horas durante las últimas 2 semanas. Marque una respuesta por cada síntoma.

¿Cómo de grave son los siguientes síntomas en su mano?	No tengo síntomas	Leve	Moderado	Severo	Muy Severo
Dolor por la noche	1	2	3	4	5
Dolor durante el día	1	2	3	4	5
Adormecimiento (pérdida de sensibilidad) durante la noche	1	2	3	4	5
Adormecimiento (pérdida de sensibilidad) durante el día	1	2	3	4	5
¿Con qué frecuencia le despertaron a usted los siguientes síntomas por la noche?	Nunca	Una vez	2 o 3 veces	4 o 5 veces	Más de 5 veces
Dolor	1	2	3	4	5
Adormecimiento (pérdida de sensibilidad)	1	2	3	4	5

## **Versión española de la escala del dolor en la palma**

¿Cuánto dolor tiene usted en la palma de la mano o en la cicatriz quirúrgica?

Ningún dolor      Muy Leve      Leve      Moderado      Fuerte      Muy Fuerte

¿Hasta qué punto el dolor en la palma de la mano o en la cicatriz quirúrgica le ha dificultado sus actividades?

Nada              Un poco              Regular              Bastante              Mucho

# 7

## BIBLIOGRAFÍA

- Alonso J, Prieto L, Anto JM. La versión española del SF-36 Health Survey (Cuestionario de Salud SF-36): un instrumento para la medida de los resultados clínicos. *Med Clin (Barc)* 1995;104:771-6
- Amadio P, Silverstein M, Ilstrup D, Schleck CD, Jensen LM. Outcome assessment for carpal tunnel surgery: the relative responsiveness of generic, arthritis-specific, disease-specific, and physical examination measures. *J Hand Surg Am* 1996; 21: 338-46
- Amadio P.C. Outcome assessment in Hand Surgery and Hand Therapy: an update. *Journal of Hand Therapy* April-June 2001.

- Arnold M. Epstein, M.D.,M.A: The Outcomes Movement-Will it get us where we want to go? . The New England Journal of Medicine vol 323. N.4 July 26, 1990.
- Atroshi Isam, MD, Warren C. Breidenbach, MD, Steven J.McCabe,MD, Louisville,KY. Assessment of the carpal tunnel outcome instrument in patients with nerve-compression symptoms. J Hand Sur 1997; 22A:222-227.
- Atroshi Isam, MD, Kristiansd, Sweden, Ragnar Johnson,MD, PhD, Lund, Swewden, Ewald Ornstein, MD, Kristianstad, Swewdn. Patient Satisfaction and Return to Work after Endoscopic Carpal Tunnel Surgery. J Hand Surg 1998; 23A:58-65.
- Atroshi Isam, MD, Christina Gummesson, MS, Ragnar Johnsson. Prevalence of carpal tunnel syndrome in a general population. JAMA, July 1999. Vol281, n°2.
- Atroshi Isam, MD, Christina Gummesson, MS, Kristianstad, Sweden, Ragnar Johnsson, MD, PhD, Lund, Sweden, Anna Sprinchorn, MD, Kristianstad, Sweden. Symptoms, disability, and quality of life in patients with carpal tunnel syndrome. J Hand Surg 1999; 24A:398-404
- Atroshi I, Lyren PE, Gummesson C. The 6-item CTS symptoms scale: a brief outcomes measure for carpal tunnel syndrome. Qual Life Res. 2009; 18: 347-58
- Atroshi I, Christina Gummesson. Non- surgical treatment in carpal tunnel syndrome. The Lancet. vol 374. September 26, 2009.

- 
- Atroshi I, Lyren PE, Ornstein E, Gummesson C. The six-item CTS symptoms scale and palmar pain scale in carpal tunnel syndrome. *J Hand Surg Am.* 2011; 36: 788-94.
  - Badalamente, PhD, Lauren Coffelt, BScot and cols. Measurements Scales in Clinical research of the upper extremity: outcome measures in studies of the hand, wrist and shoulder, elbow. *JHand Surg* 2013; 38A:407-412.
  - Beaton D.E., James G.Wright, Jeffrey N. Katz. Development of the QuickDASH: Comparison of three item-reduction approaches. *J Bone Joint Surg Am.* 2005;87:1038-1046.
  - Bentsianov BL, Boruk M, Rosendfield RM. Evidence-based medicine in otolaryngology journals. *Otolaryngol Head Neck Surg.*
  - Brent Graham, MD, Glenn Regehr, PhD, Gary Naglie, MD, James G. Wright, MD. Development and validation of diagnostic criteria for carpal tunnel syndrome. *JHand Surg* 2006; 31A:919. e1- 919 e7.
  - Brent Graham, MSC, MD. Alternatives to randomized trials in clinical research. *Journal of the American Society for Surgery of the Hand*, 2005; vol 5,nº1.
  - Cáceres-Palou E, F.López-Prats, M.Mesea-Ramos, J.Sánchez Sotelo, S.Suso-Vergara. 2005. Valoración de resultados en Cirugía Ortopédica y Traumatología. *Rev Orthop Traumatol*; 49(supl 1):119-42
  - Calfee RP, MD, Amelia A. Adams. Clinical research in patient- rated outcomes measures in hand surgery. *JHand Surg* 2012;37A:851-855

- Chatterjee Justin S, MBChB, Patricia E. Price, PhD. Comparative responsiveness of the Michigan Hand Outcomes Questionnaire and the Carpal Tunnel Questionnaire after carpal tunnel release. *J. Hand Surg* 2009; 34 A: 273-280.
- Christina Gummesson, MSPT, Lund; Swewden, Isam Atroshi, MD, PhD, Kristianstad, Swewden, Charlotte Ekdahl, PhD, Lund, Swewden: "The Quality of Reporting and Outcome Measures in Randomized Clinical Trials related to Upper-Extremity Disorders". *The Journal Hand Surgery* 2004; 29A: 727-734.
- Christina Gummesson, Michel M Ward, Isam Atroshi. The shortened disabilities of the arm, shoulder and hand questionnaire (QuickDASH): validity and reliability based on responses within the full-length DASH. *BMC Musculoskeletal disorders* 2006, 7:44.
- Donabedian A. La calidad de la asistencia. *JANO*;36:103-10.
- Gandek B, Ware JE, Aaronson Nk, Apolone G, Bjorner JB, Brazier Je, et al. Cross-validation of item selection and scoring for the SF-12 Health Survey in nine countries: results from the IQOLA Project. *International Quality of Life Assessment. J Clin Epidemiol.* 1998;51:1171-8
- Gay R,MD, Peter Amadio, MD, Rochester, MN, Jane C. Johnson, MA, Kirksville, MO. Comparative Responsiveness of the Disabilities of the arm, shoulder, and hand, the carpal tunnel questionnaire, and the SF-36 to clinical change after carpal tunnel release. *J Hand Surg* 2003; 28A:250-254.

- 
- Greenslade J.R., R.L. Mehta. P.Belward and D.J. Warwick. DASH and Boston questionnaire assessment of carpal tunnel syndrome outcome: What is the responsiveness of an outcome questionnaire? J Hand Surg (British and European Volume, 2004) 29B: 159-164.
  - Hanzlik S, Mahabir RC, Baynosa RC, Khiabani KT. Levels of evidence in research published in the journal of bone and joint surgery (American volume) over the last thirty years. J Bone Joint Surg. 2009;91A:425-8.
  - Haynes RB, Sackett DL, Tugwell P. Problems in the handling of clinical and research evidence by medical practitioners. Arch Intern Med. 1983; 143:1971-5
  - Hobby J.L. C. Watta and D. Elliot. Validity and Responsiveness of the patient evaluation measure as an outcome measure for carpal tunnel syndrome. Journal of Hand Surgery ( British and European volume, 2005) 30B 4: 350-354
  - Horton R. Surgical research or comic opera: questions, but few answers. Lancet. 1996;347:984-5.
  - Jadad AR, Moore RA, Carroll d, Jenkinson C, Reynolds DJM, Gavaghan DJ, McQuay HJ (1996): "Assessing the quality of reports of randomized clinical trials: is blinding necessary?" Controlled Clinical Trials, 17 (1): págs. 1-12; 1996.
  - Joseph Bernstein, MD,MS: Medicina basada en evidencias: J Am Acad Orthop Surg (Ed Esp) 2004;3:152-160.

- J.M.Valderas, Montse Ferrer, Jordi Alonso. Instrumentos de medida de calidad de vida relacionada con la salud y de otros resultados percibidos por los pacientes. Med Clin (Barc) 2005; 125(Supl.1):56-60
- Kevin C. Chung, MD, Matthew S. Pillsbury, BS, Madonna R. Walters, BSN, Rodney A. Hayward, MD, Ann Arbor, MI. Reliability and Validity testing of the Michigan Hand Outcomes Questionnaire. J Hand Surg 1998; 23A:575-587.
- Kevin C. Chung, MD,MS, Loree K. Kallianen, MD, Rodney A. Hayward, MD, Ann Rbor, MI. Type II ( $\beta$ ). Errors in the Hand Literature: The importance of power. J Hand Surg 1998;23A:20-25.
- Kevin C. Chung, MD, Ann Arbor, MI. Commentary: The Quality of Reporting and Outcome Measures in Randomized Clinical Trials Related to Upper Extremity Disorders. The Journal of Hand Surgery 2004: Vol 29 No. 4 July.
- Kevin C. Chung, MD,MS, Patricia B. Burns, MPH, Erika Davis Sears, MD. Outcomes research in Hand Surgery: Where have we been and where should we go? J Hand Surg 2006; 31A:1373-1379.
- Kevin C. Chung, MD. The minimal Clinically Important Difference of the Carpal Tunnel Syndrome Sympton Severity Scale. The Journal Hand Surgery. vol.31A No.5 May-June 2006
- Kevin Chung, MD. Clinical Research in Hand Surgery. 2010: J Hand Surg; 35A: 109-120.
- Keller RB 1996. Measuring outcomes. J Orthop Res. 1996;14:171-2

- 
- Landis JR, Koch GG. The measurements of observer agreement for categorical data. *Biometrics*. 1977;33:159-74.
  - LaiTY, Leung GM, WongVW, Lam RF, Cheng AC, Lam DS. How evidence-based are publications in clinical ophthalmic journals? *Invest Ophthalmol Vis Sci*.2006;47:1831-8.
  - Levine D, Simmons B P, Mark J K, et al.. A self-administered questionnaire for the assessment of symptom severity and functional status in carpal tunnel syndrome. *J Bone J Surg Am*. 1993; 75: 1585-92.
  - Lyren PE, Atroshi I. Using item response theory improved responsiveness of patient reported outcomes measures in carpal tunnel syndrome. *J Clin Epidemiol*. 2012;
  - Marc F. Swiontkowski, MD. The Outcomes Movement in Orthopaedic Surgery: Where We are and Where we should Go ? *The Journal of bone and Joint Surgery*,1999.
  - Medical Outcomes Trust Scientific Advisory Committee. Instrument review criteria. *Medical Outcomes Trust Source Pages*. 1997:6-9
  - Mokkink Lb, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, Bouter LM, deVet HC. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. *BMC Med Res Methodol*.2010;10:22.doi:10.1186/1471-2288-10-22
  - Mother D, Schulz KF, Altman DG. The CONSORT Statement: revised recommendations for improving the quality of reports of parallel group randomized trials. *BMC Med Res Methodol* 2001; 1:2

- Obremskey WT, Pappas N, Attallah-Wasif E, Tornetta P, Bhandari M. Level of evidence in orthopedic journals. *J Bone Joint Surg.* 2005;87A:2632-8.
- Oxford Center for Evidence-Based Medicine: Levels of Evidence. Disponible en: <http://www.cebm.net/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/>.
- Özyürekoglu Tuna, MD, Steven I. McCabe, MD, MSc, L. Jane Goldsmith, PhD, A.Scott LaJoie, PhD. The Minimal Clinically Important Difference of the Carpal Tunnel Syndrome Symptom Severity Scale. *J Hand Surg* 2006; 31A:733-738
- Page RM, Cole GE, Timmreck TC. *Basic Epidemiological Methods and Biostatistics. A practical guidebook.* Boston: Jones and Bartlett publishers; 1995.
- Rebozo Morales LE, Rosales RS, Dorta Fernandez A, Afonso Rguez J, Cerdan S, Diez de Lastra I. The combined use of specific instruments in the assessment of health status diagnosis in carpal tunnel syndrome. 7th Congress of The IFSSH Vancouver Canada May 24-28 1998. ISBN 88-323-0524-0 (Trabajo becado por el fondo de Investigación Sanitario)
- Robert M. Szabo, MD, Sacramento, CA. Outcomes assessment in hand surgery: when are they meaningful? *J Hand Surg* 2001; 26A: 993-1002

- 
- Robert M. Szabo, MD: "Show Me the Evidence". *J Hand Surgery* 2008;33A:150-156.
  - RS Rosales, A Dorta Fdez, L Rebozo Morales, J Afonso-Rguez, S Cerdan. I Diez de Lastra. The use of health instruments in hand Surgery. Method of adaptation and reliability of the Spanish version of most used generic and specific questionnaires. 7th Congress of the IFSSH Vancouver Canada May 24-28 1998.
  - R. S. Rosales, MD, PhD, Elena Benseny Delgado, PsyD, Isabel Díez de la Lastra-Bosch, MD. 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* 2002; 27A:334-343.
  - Rosales S, Rodríguez Pulido F. Discapacidad de MMSS y Calidad de Vida en el Músico. *Rev Iberoamer Cir Mano*, 2005; 33:75
  - Rosales. Validez de Construcción de la versión española del DASH. XV Congreso de la SECMA, 2011.
  - Rosales RS, Atroshi I. Spanish versions of the 6-item carpal tunnel syndrome symptoms scale (CTS-6) and palmar pain scale. *J Hand Surg Eur*. 2013; 38:550-1.
  - R.S. Rosales: Clinical Research in Hand Surgery. *The Journal of Hand Surgery (European Volume)* 2015 Vol.40E (5) 546-548.

- Sackett David L, William MC Rosenberg, JA Muir Gray, R Brian Haynes, W Scott Richardson. Medicina Basada en la Evidencia: Lo qué es y lo qué no es. BMJ 1996; 312: 71-2.
- Sackett DL. 1986. Rules of evidence and clinical recommendations on the use of antithrombotic agents. Chest 89(2 suppl): 25-35
- Schädel-Höpfner M, Windolf J, Antes G, et al. Evidence-based hand surgery: the role of Cochrane reviews. J Hand Surg Eur, 2008; 33: 110-7.
- Shiung-Yi Horng, MD, Ming-Chuan Lin, MD, Chi-Tzu Feng, BS, Chi-Hung Huang, BS, Hsin-Chi Wu, MD, Jung-Der Wang, MD. Responsiveness of the Michigan Hand Outcomes Questionnaire and the Disabilities of the Arm, Shoulder, and Hand questionnaire in patients with hand surgery injury. J Hand Surg 2010; 35A:430-436
- Silman AJ. Epidemiological Studies: a practical guide. New York: Cambridge University Press; 1995. Sinno H, Neel OF, Lutfy J, Bartlett G, Gilardino M. Level of evidence in plastic surgery research. Plast Reconstr Surg. 2011; 127:974-80.
- Sinno H, Neel OF, Lutfy J, Bartlett G, Gilardino M. Level of evidence in plastic surgery research. Plast Reconstr Surg. 2011; 127:974-80.
- Tadjerbashi K, Rosales RS, Atroshi I. Intervention randomized controlled trials involving wrist and shoulder arthroscopy: a systematic review. BMC Musculoskelet Disord. 2014, 25; 252.
- Terwee CB, Mokkink Lb, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating methodological quality in systematic reviews of studies on measurement

properties: a scoring system for the COSMIN checklist. *Qual Life Res.* 2012;21:651-57

- Valderas J.M, Montse Ferrer y Jordi Alonso. Instrumentos de medida de calidad de vida relacionada con la salud y de otros resultados percibidos por los pacientes. *Med Clin (Barc)*, 2005; 125(Supl. 1): 56-60
- Ware JE, Snow KK, Kosiski M, Gandek B. *SF-36 Health Survey: manual and interpretation guide*. Boston: New England Medical Center; 1993