

ARTÍCULO SOLICITADO

Metodología para el desarrollo de un trabajo de investigación clínica en Cirugía de la Mano

R.S. ROSALES⁽¹⁾, F. RODRÍGUEZ PULIDO⁽²⁾, I. ATROSHI⁽³⁾, Y. MARTÍN HIDALGO⁽¹⁾, S.J. McCABE⁽⁴⁾

⁽¹⁾ UNIDAD DE CIRUGÍA DE LA MANO Y MICROCIRUGÍA. GECOT. TENERIFE. ESPAÑA

⁽²⁾ UNIDAD DOCENTE DE PSIQUIATRÍA. DEPARTAMENTO DE MEDICINA INTERNA.

UNIVERSIDAD DE LA LAGUNA. TENERIFE. ESPAÑA

⁽³⁾ DEPARTMENT OF ORTHOPEDICS, LUND UNIVERSITY, LUND; HASSLEHOLM-KRISTIANSTAD HOSPITALS.

SUECIA

⁽⁴⁾ DEPARTMENT OF BIOINFORMATICS, SCHOOL OF PUBLIC HEALTH AND INFORMATION SCIENCES.

THE UNIVERSITY OF LOUISVILLE. LOUISVILLE, KENTUCKY, USA

Correspondencia:

Dr. Roberto S. Rosales

Unidad de Cirugía de La Mano y Microcirugía. GECOT

María del Cristo Osuna 20 bajos

38204 La Laguna. Tenerife

Teléfono 922 251 233

Fax 922 251 541

e- mail: titorosales@telefonica.net

El propósito del presente trabajo es describir las pasos principales en el desarrollo de un trabajo de investigación clínica en cirugía de la mano, con el objetivo de establecer una guía para aquellos investigadores y cirujanos de mano interesados en el campo de la valoración de resultados. Así se aborda: la importancia de los antecedentes en la generación de la pregunta de investigación y los objetivos en función de los diferentes tipos de trabajos de investigación clínica (*el porqué*); el material y método (*el cómo*), introduciendo los principales apartados que debe incluir, como son la definición de la población estudio, diseño clínico, instrumentos y mediciones, y análisis estadístico; la descripción de los resultados (*lo descubierto*); para finalizar con la discusión de los mismos (*la interpretación de los resultados*).

Palabras clave: metodología, investigación clínica, resultados, DASH.

The purpose of this review paper is to describe the fundamental steps for the developing of a clinical research paper in hand surgery with the aim to establish a guideline for young researchers and hand surgeons interested on the outcomes research field. The paper describes: the generation of the research question and the purpose of the different clinical research papers (*Why?*); the material and method (*How?*) showing the sections which should be included: study population, clinical design, instruments and measures, and data analysis; the results (*what has been discovered?*); and, finally, the discussion (*interpretation of results*)

Key Word: methodology, clinical research, outcomes, DASH.

Los avances tecnológicos han permitido que la Cirugía de la Mano alcance límites insospechados. Hace algo más de 40 años se realizaba por primera vez una sutura arterial digital¹, y hoy se realizan, de manera sistemática, colgajos libres vascularizados, injertos nerviosos, transferencias de dedo de pie a mano, e incluso trasplantes de mano de cadáver. Pero este desarrollo técnico no se ha acompañado de una mejora en la calidad de los trabajos de investigación clínica publicados: se sigue confundiendo incidencia con prevalencia cuando se habla de la ocurrencia de una enfermedad o patología crónica en la mano²⁻⁵; no se definen correctamente las poblaciones dianas o estudio; se sigue utilizando la palabra *retrospectivo* en el título de un estudio, cuando tanto un diseño clínico descriptivo de corte transverso, como un estudio de casos control, son retrospectivos, pero con niveles de evidencia científica diferentes⁶. El nivel de los ensayos clínicos y estudios observacionales, publicados en las principales revistas científicas relacionadas con la mano y miembros superiores (MMSS), es muy bajo y de muy poca calidad;^{7,8} lo que repercute a la hora de la revisión sistemática de la evidencia científica de los trabajos publicados, haciendo casi imposible la realización de meta-análisis y el establecimiento de guías clínicas y protocolos de actuación ante patologías importantes en la Cirugía de la Mano⁹.

El propósito del presente trabajo es describir los pasos principales en el desarrollo de un trabajo de investigación clínica en Cirugía de la Mano, con el objetivo de establecer una guía para aquellos investigadores y cirujanos de mano interesados en la metodología de la valoración de resultados. Así se aborda: la importancia de los antecedentes en la generación de la pregunta de investigación u objetivos, en función de los diferentes tipos de trabajos de investigación clínica (*el porqué*); el material y método (*el cómo*), introduciendo los principales apartados que debe incluir como la definición de la población estudio, diseño clínico, instrumentos y mediciones, y análisis de los datos; la descripción de los resultados (*lo descubrierto*); para finalizar con la discusión de los mismos (*la interpretación de los resultados*).

En los siguientes apartados se presentan normas generales de referencia para los lectores, investigadores y cirujanos interesados, que deben completarse con los principios metodológicos de preparación de cualquier proyecto de investigación¹⁰⁻¹².

INTRODUCCIÓN

En esta sección del trabajo se analizan los antecedentes y el objetivo del trabajo de investigación.

Los antecedentes y la bibliografía de cualquier proyecto de investigación clínica permiten determinar la familiaridad de los autores de un trabajo con el tema objetivo del estudio, así como juzgar la importancia o necesidad del estudio propuesto.

La sección de introducción, o de antecedentes, es denominada en las revistas anglosajonas de mayor impacto «background». En esta sección se debe exponer cuál es el conocimiento científico actual sobre el tema objetivo del estudio (*qué se conoce*), cual es el salto en el conocimiento (*qué no se conoce*), generar la pregunta de investigación e hipótesis del trabajo, 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 que el investigador pretende resolver o contestar mediante la realización de mediciones, siguiendo un método, y en una serie de sujetos que constituye la población estudio. Así, por ejemplo, McCabe et al.¹³, mediante un proceso denominado de «destilación», establecen nuevas preguntas sobre la relación causal entre la posición de la mano y disturbios del sueño (ej. apnea, roncar, etc.) con la obesidad y el STC. La pregunta de investigación debe ser: importante (por la gravedad, frecuencia, etc.); interesante (a la luz de los conocimientos actuales); que tenga respuesta (con los medios disponibles), y realizables (en un tiempo concreto).

Según Haynes et al.¹⁴, en función de la cuestión fundamental que el investigador quiere contestar con su trabajo, los estudios de investiga-

ción clínica se pueden clasificar en: estudios sobre un test o pruebas diagnósticas; estudios sobre el pronóstico de una enfermedad; estudios sobre la etiología o causalidad; estudios sobre la eficacia-efectividad de una terapia o tratamiento (outcomes) (**Figura 1**).

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 generación 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».

Estudios sobre el pronóstico de una enfermedad, son aquellos destinados a conocer el curso clínico de un proceso patológico y en ellos se debe revisar si los estudios previos están basados en un correcto diseño tipo cohortes, que será abordado en el apartado de diseños clínicos del presente trabajo.

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. Como se verá más adelante, en este tipo de estudios es importante diferenciar entre dos tipos de diseños clínicos establecidos para demostrar asociación causal con una enfermedad, y que son: los estudios de casos-control y los cohortes.

Estudios sobre la eficacia y efectividad de un tratamiento medicamentoso o quirúrgico. Estos estudios de investigación 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. En ocasiones 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 ensayo clínico aleatorizado. Hablaremos de efectividad cuando se haya medido el resultado, el efecto, de esa decisión en lo que se conoce como condiciones habituales, práctica clínica habitual, y eso se mide mediante estudios observacionales. Y entenderemos como eficiencia la medida del coste de oportunidad de la decisión en sanidad, es decir, si se invierte aquí en que se deja de invertir, y si haber dejado de invertir en un recurso es más beneficioso para esa población que hacer lo contrario. Los ensayos o los análisis de coste-efectividad serían el método para determinar la eficiencia de una determinada decisión¹⁵. Un tratamiento puede ser sumamente eficaz en mejorar a los pacientes pero poco coste-efectivo. Normalmente, los estudios de coste-efectividad se establecen para distinguir, para una patología determinada, de entre aquellos tratamientos considerados buenos (de probada eficacia) cual 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). Así por ejemplo, Chung et al.¹⁶, compararon el coste-efectividad de la liberación endoscópica del túnel carpiano (LETG) con la liberación estándar abierta (LATG), demostrando un ratio de efectividad de 195\$/QALY y de 693\$/QALY respectivamente, pudiéndose comparar incluso con otras intervenciones sanitarias, como el cribado de cáncer de mama 4836\$/QALY, o con el uso del ejercicio físico para prevenir la enfermedad coronaria 13,508 \$/QALY.

La sección de introducción debe redactarse en tiempo pasado, de forma precisa y detallada, aunque no enciclopédica. Tanto la sección de introducción como la bibliografía seleccionada deben estar orientadas hacia la hipótesis de investigación. En la sección de «introducción» de cualquier trabajo de investigación clínica los autores deben demostrar el conocimiento de los resultados obtenidos previamente en ese campo, deben saber qué estudios se están realizando, cuáles son sus posibles limitaciones, y qué preguntas quedan por responder. La descripción de generalidades en la sección de «introducción», las referencias bibliográficas obsoletas y genéricas, y la falta de concreción al

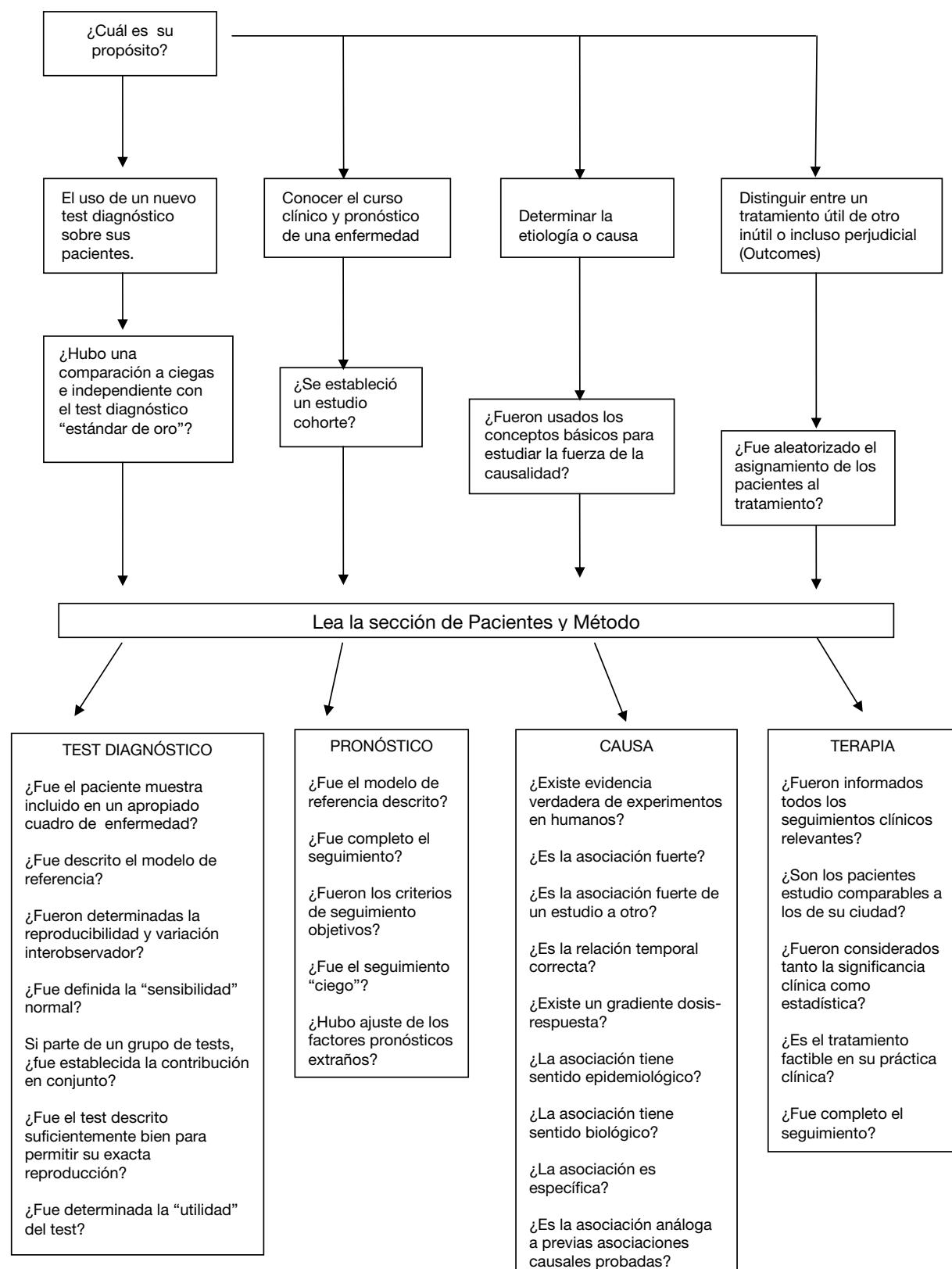


Figura 1. Tipos de estudios de investigación clínica en base a la pregunta de investigación. Obsérvese los puntos a revisar en el material y método de cada tipo de trabajo de investigación a la hora de realizar la revisión bibliográfica y establecer la pregunta de investigación.

detallar los problemas que quedan por resolver, pueden indicar que los investigadores no están excesivamente familiarizados con ese campo de investigación, lo que no indica falta de competencia en la práctica de esa disciplina.

La hipótesis sintetiza la pregunta científica a la que se pretende dar respuesta con el trabajo de investigación.

El objetivo/s, o propósito del trabajo, constituye la declaración explícita de lo que se pretende alcanzar con el estudio. Deben ser concretos, evaluables, factibles, congruentes y deben tener una clara relevancia científica o socio-sanitaria. Objetivos excesivos, incongruentes o mal definidos indican un mal planteamiento del proyecto en el que se basó el estudio de investigación clínica. Finalmente, el número de objetivos de un trabajo debe ser limitado, aunque proyectos complejos (multidisciplinarios o multicéntricos) pueden presentar objetivos más ambiciosos. En ese caso, se debe priorizar en orden de importancia, distinguiéndose entre objetivo/s primario/s (que responden a la pregunta principal de investigación) y secundarios (aspectos parciales de la investigación o respuestas derivadas de los objetivos principales).

MATERIAL Y MÉTODO

La sección de Material y Método constituye la parte más importante de cualquier trabajo de investigación clínica. De ella dependerá si los resultados obtenidos pueden aceptarse como evidencia científica. La metodología de un trabajo clínico debe describir paso a paso, y con el suficiente grado de detalle, las acciones que se llevaron a cabo para alcanzar los objetivos. Los autores tienen la responsabilidad de ser explícitos al presentar los detalles de su propuesta, de forma que los lectores tengan que realizar el mínimo de suposiciones o de juicios de valor sobre la adecuación de la metodología. Aunque en determinados proyectos la metodología puede agruparse en otros apartados, en general, cualquier trabajo de investigación clínica debe incluir en la sección de Material y Método los siguientes apartados: Definición de la

Población Estudio, Diseño Clínico, Instrumentos y Mediciones, y Análisis Estadístico¹⁰; descritos de manera que el estudio pueda ser reproducible.

A. Definición de la Población Estudio

En la metodología de los proyectos de investigación uno de los aspectos de especial relevancia es el relativo a la población de estudio. En éste debe detallarse la población de referencia de la que proceden los pacientes y las muestras. Es importante, asimismo, detallar criterios de inclusión/exclusión concretos y operativos, así como justificar el tamaño muestral que se propone incluir en el estudio. El investigador debe tener en cuenta que los pacientes que va a estudiar serán tomados de una población particular. Por consiguiente, es importante que pueda definir las características de esa población. Debe poder dar una información acerca de las características socio-demográficas, clínica y social de la muestra. Más específicamente, debe tener conocimiento sobre cualquier aspecto o característica de la población sobre la que se toma la muestra que pueda influir de algún modo en el resultado del estudio, de acuerdo a la revisión previamente realizada. Los criterios de inclusión definen las características de la población elegible para el estudio y vienen determinados por la propia pregunta que intenta resolver la investigación. Supongamos que un investigador quiere valorar la eficacia de una nueva técnica quirúrgica para el tratamiento del síndrome del túnel carpiano (STC), ¿Qué criterios debería emplear?. La definición de un paciente como «caso» y su inclusión en la población estudio dependerá de los criterios diagnósticos que empleemos para asegurarnos que vamos a medir resultados en pacientes afectos de STC. Así llegamos a 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 del cualquier diagnóstico médico va a depender de atributos específicos: sensibilidad (capacidad de una prueba

Tabla I - RESULTADOS DE LA PRUEBA DIAGNÓSTICA Y LA EXISTENCIA DE LA ENFERMEDAD

		Criterio de verdad				
		Prueba de referencia o «estándar de oro»				
Prueba diagnóstica	Positivos	Enfermos	No enfermos	Total		
	Negativos	VP	FP	VP + FP		
	Total	FN	VN	FN + VN		
		VP + FN	FP + VN	N		
				(VP + FP + FN + VN)		

VP: verdadero positivo; FP: falso positivo; FN: falso negativo; VN: verdadero negativo.

diagnóstica en detectar los verdaderos positivos (VP) entre todos los que verdaderamente están afectos de la enfermedad: VP/VP+FN (FN: Falsos Negativos), especificidad (capacidad de una prueba diagnóstica en detectar los verdaderos negativos (VN) de entre todos los pacientes que no están afectos de la enfermedad: VN/VN+FP (FP: Falsos Positivos)¹⁷ (**Tabla I**). Cuanto más sensible sea una prueba más probable es que detecte a las personas o sujetos con la enfermedad. Por ello, las pruebas o criterios con sensibilidad elevada son muy útiles clínicamente para descartar la presencia de la enfermedad o condición patológica bajo estudio. Es decir, con una prueba así, un resultado negativo casi excluiría la posibilidad de que el paciente tenga la enfermedad. Esto se deriva del hecho de que una prueba muy sensible provoca pocos FN.

En cuanto a la especificidad, cuanto una prueba es más específica, 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. Si la prueba es muy específica, y da positiva, nos está indicando la presencia de la enfermedad, ya que apenas produce FP. Finalmente, necesitamos conocer qué nivel de sensibilidad y especificidad deberíamos usar en una prueba diagnóstica o criterio para definir a un paciente como «caso». Para ello se utilizan las «Curvas ROC», desarrolladas originalmente por fí-

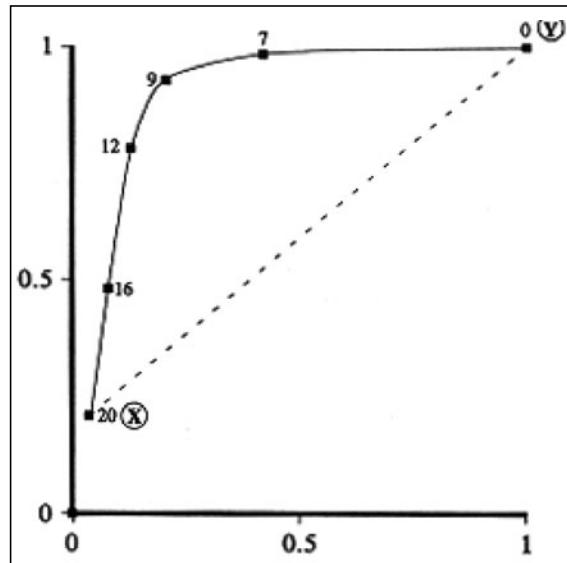


Figura 2. Ejemplo de curva ROC. Resultados de un cuestionario para determinar la presencia de neuropatía compresiva en MMS. El cuestionario de 20 ítems es puntuado de 0 a 20. Los valores de sensibilidad y especificidad son calculados de una muestra de 200 individuos, todos fueron entrevistados con el cuestionario y además fueron clínicamente evaluados. Con un punto de corte X (equivalente a 20), muy pocos individuos serán definidos como «caso» (sensibilidad de 0,2) pero a la vez presentarán muy pocos falsos positivos (especificidad de 0,95). Con un punto de corte Y (equivalente a 0) todos los individuos serán definidos como caso (sensibilidad = 1) pero todos los verdaderos negativos (VN) clasificados como positivos (especificidad = 0). El máximo de sensibilidad y especificidad lo da el ápex de la curva (punto de corte = 8). Un corte por encima (por ejemplo de 12), daría muchos verdaderos positivos, y un corte por debajo (por ejemplo de 7), daría muchos falsos positivos.

sicos ingleses, que, en la segunda guerra mundial, estudiaban las señales de radar para distinguir las señales de bombas volantes de los alemanes de otro tipo de alteraciones del radar, y de ahí sus siglas (ROC: Received Operating Characteristic), extrañas en el campo de la medicina, por eso a veces son denominadas curvas de rendimiento diagnóstico. Las curvas de ROC son gráficos en los que se enfrentan en el eje de ordenadas la Sensibilidad (S), y en el eje de abcisas el complemento de la Especificidad ($1 - E$), o lo que es lo mismo los FP (**Figura 2**). Por ejemplo, tenemos un cuestionario de 20 preguntas para detectar mediante encuestadas en la población general aquellos posibles sujetos afectos de neuropatía compresiva en manos y MMSS, y quisiéramos saber cuantas preguntas deben ser contestadas para definir al encuestado como posible «caso». El cuestionario de 20 ítems puede ser puntuado de 0 a 20. Como se observa en la figura si adoptamos un punto de corte «X» (cut-off point) equivalente a 20 muy pocos individuos serán definidos como «caso» (baja sensibilidad $S = 0,2$), pero a la vez presentará muy pocos falsos positivos (alta especificidad $E = 0,95$). En el extremo opuesto, si tomamos como punto de corte a «Y» equivalente a cero, todos los individuos encuestados serán definidos como «caso» lo que significa muy alta sensibilidad ($S = 1$) pero con todos los VN clasificados como positivos (100% FP) ($1-E = 1$), o lo que es lo mismo especificidad igual a cero. El nivel de puntuación correcto lo da el ápex de la curva (en este ejemplo punto de corte = 8) donde obtenemos el máximo de sensibilidad y especificidad. El área que queda debajo de la curva hasta la diagonal, es una medida global de la exactitud de la prueba y se define como la probabilidad de clasificar correctamente un par de individuos, sano y enfermo, seleccionados al azar de la población, mediante los resultados obtenidos al aplicarles la prueba diagnóstica o criterio de definición¹⁷. Las curvas ROC han sido utilizadas en cirugía de la mano para determinar la sensibilidad y especificidad de diferentes test clínicos y neurofisiológicos en el diagnóstico de STC^{18, 19}, para detectar, en la población general, aquellos STC graves que potencialmente necesitan descompresión quirú-

gica mediante una encuesta con cuestionarios²⁰; para conocer que existe un 0,7 % de la población general que está afecta de STC grave y que no ha recibido tratamiento²⁰; para calcular la diferencia mínima de importancia clínica que se puede medir a nivel de la gravedad de síntomas en el STC²¹; para analizar la capacidad de los instrumentos de medición DASH y del QuickDASH en discriminar entre pacientes clasificados en base a la magnitud de la mejoría clínica observada²²; o para calcular la precisión y exactitud diagnóstica de las radiografías (Rx), simples y dinámicas, en el diagnóstico de la disociación escafo-semilunar²³.

Una vez introducido, la importancia de los atributos de sensibilidad y especificidad a la hora de elegir los criterios de inclusión para definir una población estudio, podemos concluir que 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. Por ejemplo, no será la misma población estudio de STC cuando el investigador quiere analizar la eficacia del tratamiento en STC graves, en pacientes de edad avanzada, que en trabajadores manuales. El investigador define sus criterios y eso constituye siempre una variable que es controlada por el investigador. El número de criterios va a depender de la decisión del investigador, y para ello siempre debe consultar la evidencia publicada con anterioridad. Así, Atroshi et al.²⁴ demostraron que al analizar la prevalencia de STC en la población general, ésta dependería de los criterios diagnósticos establecidos para definir a un paciente como «caso afecto». Así, si basamos el diagnóstico sólo en los síntomas que refieren los pacientes, diagnosticaremos un 14% de afectos en la población general. Si estos casos sintomáticos eran explorados por un cirujano de mano entrenado y con experiencia clínica, la prevalencia disminuía a un 3,8%. Finalmente, si además se sumaba un estudio neurofisiológico (EN), la prevalencia disminuía a un 2,7%. Por otro lado, demostró que si el mismo EN, realizado por un neurofisiólogo independiente, y a ciegas con la clínica, se llevaba a cabo en una muestra aleatoria de los asintomáticos en la población general, un 18% de los

mismos presentaban criterios positivos neurofisiológicos de STC. Por tanto, la conclusión de este estudio es que el uso único del EN tiene un 18% de falsos positivos, y que para establecer el diagnóstico es necesario tener en cuenta tanto la clínica, como la exploración y el EN. Siguiendo estos criterios diagnosticaremos menos STC, pero estaremos seguros de que son realmente STC.

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. Por ejemplo, supóngase que un cirujano de mano quiere valorar la eficacia de un nuevo sistema de liberación endoscópica para el STC; a la hora de definir su población estudio, no sólo debe asegurarse que estamos midiendo resultados en una población STC, sino que, además, debe estar seguro que no existen factores de co-morbilidad en su muestra que pueda influir en sus resultados, como la existencia de asociación con otras neuropatías proximales (Síndrome de Doble Compresión)²⁵, o la asociación con diabetes, en la que puede existir una polineuropatía de base, o con artritis u otras enfermedades metabólicas. Todos estos posibles factores constituyen criterios de exclusión a la hora de definir la población estudio.

A la hora de presentar dichos criterios en nuestro trabajo debemos tener en cuenta ciertas recomendaciones:

1.- Debe ser enunciada con claridad la forma en la que se seleccionaron los pacientes para el estudio. El autor debe ser capaz de indicar cuáles son los criterios que ha utilizado para incluir/excluir pacientes que sean o no adecuados para su estudio. Lo preferible es que sean utilizadas nosologías diagnosticas internacionalmente aceptadas para las enfermedades, como la décima revisión de la Clasificación Internacional de Enfermedades (CIE-10) de la Organización Mundial de la Salud (OMS)²⁶, o si se valora el funcionamiento y/o la discapacidad, la Clasificación Internacional del Funcionamiento de la Discapacidad y de la Salud (CIF, 2001)²⁷. Estas nosologías están construidas con criterios operativos, y si en ellas el autor no en-

contrara el consenso de experto adecuado debido a que va a estudiar una patología infrecuente o rara, entonces debe remitirse a los consensos establecidos por las Sociedades Científicas Internacionales sobre esta materia. Es aconsejable utilizar definiciones ampliamente aceptadas, utilizadas y validadas en estudios previos, que van a permitir la comparación de los resultados. Puede ocurrir que al definir operativamente determinadas variables, como los criterios de inclusión y exclusión, el investigador tenga que alejarse de la definición conceptual, es decir, de lo que le gustaría medir, y tenga que asumir lo que puede medir, siendo consciente de que esta midiendo una variable o constructor sólo aproximado. Ocurre en variables que admiten diferentes conceptualizaciones y por tanto no se ha llegado a un consenso a la hora de definirlas.

2.- En el caso de que existan Entrevistas Diagnósticas Estructuradas o Semi-estructuradas, como ocurre en la actualidad en el ámbito de la atención especializada, debe recordarse que éstas deben ser administradas por personal especializado y debidamente entrenado en el uso de estos instrumentos. Recuerde que las definiciones de «caso probable» no sólo deben ser conceptuales sino también operativas, es decir, no solamente hay que tener en cuenta lo que se va a obtener con las mediciones, sino también que cantidad y en qué medida. Por lo tanto los criterios de inclusión dirán claramente «quién o qué está dentro» y los criterios de exclusión especificarán «quién o qué está fuera». La precisión, la especificación y el realismo, son tres componentes esenciales en los criterios de inclusión y exclusión. Debe establecerse criterios realistas que permitan la inclusión del número deseado de sujetos en el tiempo previsto. Es muy frecuente que se sobreestime el número potencial de candidatos; y los sujetos sometidos a estudio no deben ser muy diferentes de la población diana a la que se desearía generalizar los resultados. No debe pues omitirse ningún dato relevante, sólo la revisión bibliográfica nos dará las variables que pueden influir en los resultados.

3.- En lo que se refiere al estado clínico de los pacientes, el investigador debe ser capaz de proporcionar información tal como qué sín-

tomas o signos subjetivos deben estar presentes en el momento del estudio, y si es necesario que estén todos o un grupo de ellos, así como la intensidad y duración de los mismos (último año, última semana...). Asimismo, puede ser relevante si se está estudiando el primer episodio, o la historia de varios episodios previos, así como, el impacto de la enfermedad en el funcionamiento humano y /o cualquier tratamiento previo y actual. Así como cualquier aspecto relevante de su salud física, o mental, que pueda influir en el resultado del estudio. O por ejemplo, querer excluir pacientes que estuvieran en tratamiento previo a nuestra observación o intervención. O se puede excluir aquellos que estén enfermos o mujeres embarazadas.

4.- En la fase de planificación, el autor ya debe considerar si necesita un grupo de control, es decir, un grupo que es similar en todos los aspectos relevantes al grupo experimental con excepción de que los pacientes en el grupo de control no sufren las mismas condiciones experimentales que los pacientes del grupo experimental. En general, los grupos de control nos permiten estar suficientemente seguro de cualquier cambio que detectemos en el grupo resultado del grupo experimental, sea debido al tratamiento o técnica experimental (una nueva droga o una técnica quirúrgica) más que a alguna otra variable.

B. Diseño Clínico

Este apartado constituye la piedra fundamental del material y método, ya que del diseño clínico dependerá el nivel de evidencia científica de nuestros resultados. Este hecho es frecuentemente olvidado por los autores e incluso por los revisores imparciales (peer review, referee) en revistas científicas. Todo ello ha llevado a que revistas como el *Journal of Hand Surgery American* exija a los autores de un manuscrito que a la hora de remitirlo al editor sea clasificado en función del nivel de evidencia científica. La principal clasificación de la evidencia científica (**Tabla II**) se puede obtener en su versión original a través del Oxford Centre for Evidence Based Medicine²⁸⁻³¹ (<http://www.eboncall.org/content/levels.html>). El *Journal of Hand*

Surgery facilita una clasificación simplificada en la sección de «normas para los autores» de su web: (<http://journals.elsevierhealth.com/periodicals/yjhsu>).

Los diseños clínicos, según el nivel de evidencia científica, los podemos clasificar de la forma más sencilla 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) (**Tabla II**).

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, como las *series de casos*. Es importante destacar que todos tiene un denominador común: suelen ser retrospectivos, y sólo se incluye una medición en el tiempo del estudio. Los estudios descriptivos 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 y en general en toda la cirugía; o de grupo (estudios ecológicos) de muy poco uso en cirugía de la mano. La importancia de las *series de casos* es que pueden ser el detonante para otros estudios de mayor evidencia. Así por ejemplo, Chow y Hantes³³ hacen una muy completa revisión de 2675 casos de liberación endoscópica de STC (LETG). En ella, se describe las complicaciones observadas, y el autor concluye que la técnica presentaba un 95% de éxitos, algo que nunca se puede concluir a partir de un estudio descriptivo. De hecho, un estudio posterior publicado por Atroshi et al.³³ en el *British Medical Journal*, con un nivel superior de evidencia, demostraba mediante un ensayo clínico controlado (ECC) que no existía diferencia, en cuanto a la mejoría clínica, entre la liberación abierta (LATC) y la LETG³³.

Los *estudios de corte o transversales* son un procedimiento de observación sin seguimiento clínico (transversal), en el que una muestra re-

Tabla II - NIVELES DE EVIDENCIA CIENTÍFICA SEGÚN EL TIPO DE ESTUDIO DE INVESTIGACIÓN CLÍNICA

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 de Nivel 1; GPC† con 1b estudios de diferentes centros clínicos	RS (con homogeneidad*) de estudios de cohorte prospectivos	RS (con homogeneidad*) de estudios económicos de Nivel 1
1b	ECA individual (con Intervalo de Confianza ‡ estrecho)	Estudios de Cohorte 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***	Analisis basados en los costes clínicos o en sus alternativas; revisiones sistemáticas de la evidencia; e inclusión de análisis de sensibilidad
1c	Todos o Ninguno §	Todos o Ninguna serie de casos	Absoluta Especificidad y Sensibilidad ††	Todos o Ninguna serie de casos	Analisis del mejor valor o del peor valor absoluto †††
2a	RS (con homogeneidad*) de estudios de cohortes	RS (con homogeneidad*) de estudios de cohorte retrospectiva 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 cohortes individual (incluyendo ECA de baja calidad <80% de seguimiento)	Estudio de cohorte retrospectiva o seguimiento de controles no tratados en un ECA; derivación de GPC	Estudios de cohorte 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	Analisis 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	Analisis sin medidas de coste precisas pero incluyendo un análisis de sensibilidad que incorpora variaciones clínicamente sensibles en las variables importantes.
4	Serie de casos (estudios de cohortes y casos y controles de baja calidad §§)	Serie de casos (estudio de cohortes y pronóstico de poca calidad **)	Estudio caso-control, con pobre o no independiente estandar de referencia	Serie de casos	Analisis 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

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

Con homogeneidad, es decir, que incluye 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 significativa, y no toda heterogeneidad errónea es estadísticamente significativa. Como se comenta arriba, estudios heterogéneos deben ser señalados con un signo «» al final de su nivel designado.

Guía Práctica Clínica. (Son algoritmos o sistemas de puntuación que conducen a una estimación pronóstica o a una categoría diagnóstica.)

† Ver nota #2 para entender el uso de otros estudios con intervalos de confianza amplios

‡ 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 recoger 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, 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 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 (e.g.: usando un análisis de regresión) para encontrar qué factores son significativos. Por estudios de cohorte 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 >80%, con un tiempo adecuado para nuevos diagnósticos alternativos (e.g.: agudo 1-6 meses, crónico 1-5 años).

Tabla III - MEDIDAS APROPIADAS DE LA OCURRENCIA DE LA ENFERMEDAD

Características de la enfermedad	Ejemplos	Medidas apropiadas
A Comienzo definido claro Episodio aislado Finalizado por muerte, resolución espontánea o terapia.	Apendicitis aguda Cáncer de colon Trauma mayor	Incidencia nueva Incidencia acumulada
B Comienzo definido claro Episodios múltiples, pero infrecuentes. Corta duración Episodios terminados por muerte, resolución espontánea o terapia.	Infarto de miocardio Gripe Fracturas	Incidencia de episodio Incidencia acumulada Incidencia nueva
C Comienzo definido claro Estado de enfermedad estable crónico o que requiere una larga terapia.	Diabetes insulino-dependiente Fallo renal	Incidencia nueva Punto de prevalencia Incidencia acumulada
D Comienzo definido claro Episodios aislados o múltiples Duración variable	Artritis reumatoide Úlcera péptica	Incidencia nueva Incidencia acumulada
E Comienzo de enfermedad definido Estado de enfermedad estable crónico o que requiere una larga terapia	Hipertensión Artrosis de cadera Sordera	Punto de prevalencia
F Comienzo de enfermedad definido Múltiples episodios con períodos ausentes de enfermedad	Asma Migraña	Periodo de prevalencia Episodio de incidencia Prevalencia acumulada

presentativa 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. Su carácter transversal viene determinado porque el proceso en estudio (una enfermedad, un factor determinante de salud, etc.) se mide en un momento del tiempo. El ejemplo más claro que hay es el censo, donde toda la población se valora en un día concreto.

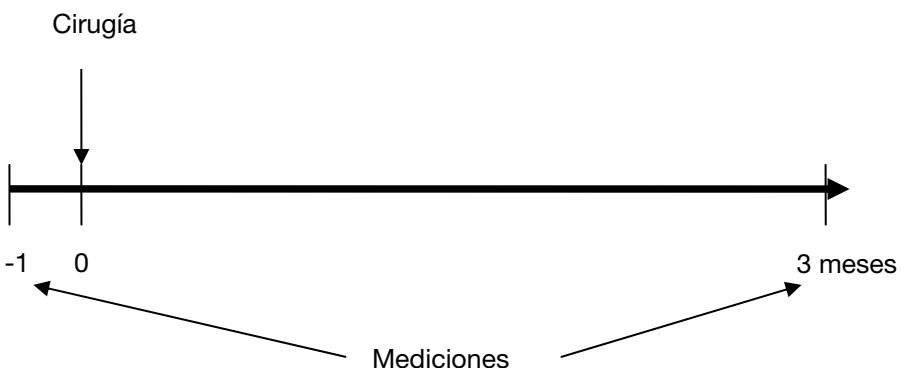
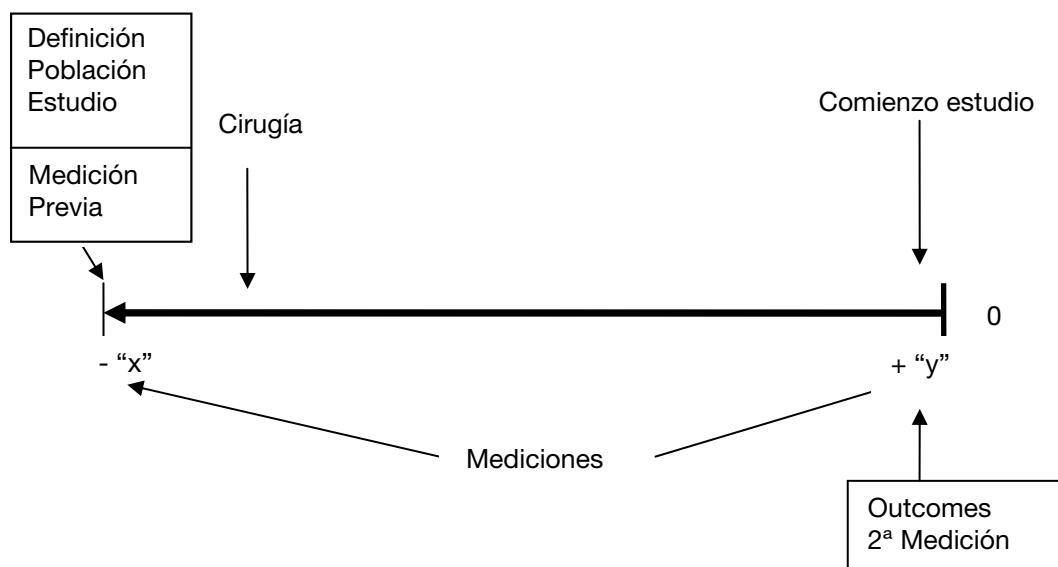
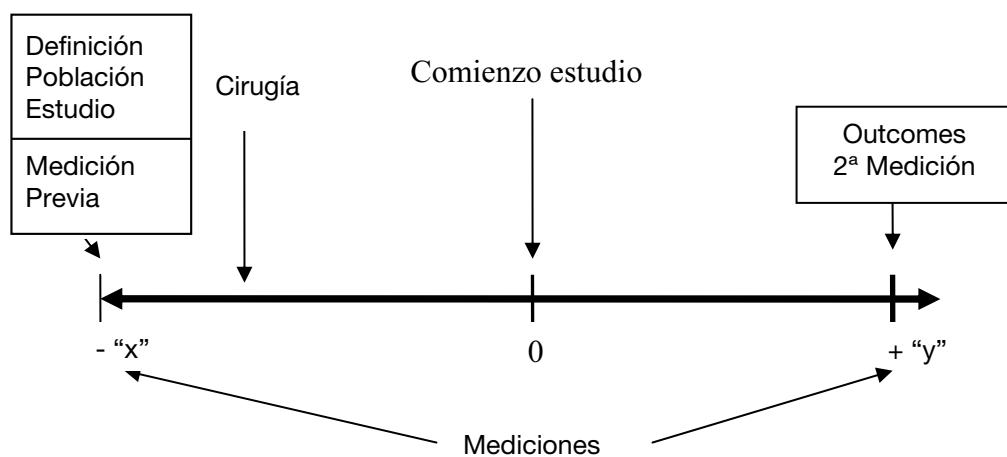
El uso de los estudios transversales se centra fundamentalmente en el diagnóstico de salud de una comunidad; en la valoración de pruebas diagnósticas; en el análisis de validez

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 fácilmente generalizables al basarse en población general³⁴. Ejemplos de estudios de corte transverso son: la valoración de la calidad de vida y discapacidad de MMSS en músicos comparada con la población general³⁵, en donde se realiza un diagnóstico de salud en la población de músicos y se compara con la población general; la comparación de una población afecta de STC y otra no afecta para

analizar la sensibilidad y especificidad de diferentes pruebas diagnósticas¹⁸; el análisis de la concordancia en el diagnóstico de STC entre la neurofisiología clásica y la portátil³⁶; la valoración de la validez del cuestionario DASH para medir resultados en una población de STC mediante la correlación con otro instrumento como el SF-12³⁷; y el análisis de la relación causal del uso del teclado y ratón del ordenador con el STC³⁸. De todos los diferentes usos de los estudios de corte o transversales antes expuestos, el más demostrativo es el análisis de la prevalencia de una enfermedad o condición patológica. Frecuentemente esta medida de la ocurrencia de la enfermedad se confunde con la incidencia²⁻⁵. La incidencia es la medida del comienzo de la enfermedad, mientras que la prevalencia es la medida del estado de la enfermedad. En la medición de la incidencia se incluyen en el numerador sólo aquellos individuos que presentan su primer episodio de la enfermedad durante un tiempo particular o período, el denominador está en relación a una población específica en la que se incluye el tiempo. Por ejemplo para la incidencia por 1000 de población por año, se necesita un estudio de seguimiento para valorar los nuevos casos. La incidencia se debe reservar para medir la ocurrencia de enfermedades con un comienzo claro, bien definido y que suelen terminar con la muerte o la resolución espontánea, o tras terapia. Así ejemplos de condiciones patológicas para medir con incidencia son: las fracturas, cáncer, apendicitis aguda, el infarto de miocardio, los brotes de gripe, etc. A diferencia, la medida de la prevalencia incluye en el numerador tanto los casos nuevos como los antiguos, en un determinado punto del tiempo (prevalencia puntual) o período de tiempo. Normalmente la prevalencia debe usarse para medir la ocurrencia de enfermedades crónicas, con un comienzo poco definido, que pueden presentar más de un episodio por año, y que suelen llevar largos períodos de tratamiento. Así, en patologías como la rizartrosis del pulgar, la seudoartrosis de escafoides, la enfermedad de Kienböck, la enfermedad de Dupuytren, y otras patologías crónicas, la medida correcta de su ocurrencia es la prevalencia¹⁷ (**Tabla III**).

Estudios Observacionales. Este grupo de diseños cínicos constituyen un nivel superior de evidencia científicas. Se distinguen los estudios de casos y controles, y los estudios de cohortes.

Estudios de casos y controles. Normalmente son diseños clínicos de carácter retrospectivo y que se dedican sobre todo al análisis causal, y la relación de factores de riesgo con el desarrollo de una enfermedad. Constituye un diseño clínico muy poco extendido en cirugía de la mano, como el publicado por Stallings et al.³⁹ en el que demuestran la asociación de la obesidad como factor de riesgo del STC, en una población de 600 pacientes que acuden a su unidad para una evaluación médica independiente. Este tipo de diseño es retrospectivo y en ocasiones puede ser confundido con otros diseños clínicos como los cohortes históricos. De ahí la importancia de no usar la palabra «retrospectivo» en el título de cualquier artículo⁶ o trabajo científico pues no explica nada sobre el diseño clínico. Así, un estudio descriptivo, un estudio de casos control o un estudio de cohorte histórico son todos retrospectivos, y tienen un nivel de evidencia científica diferente. La diferencia fundamental entre los estudios casos-controles y los cohortes está a nivel de como se seleccionan los casos o pacientes en los estudios. En los casos-controles los pacientes entran en el estudio en base a la variable resultado o dependiente (tener o no la enfermedad). Por ejemplo, si investigamos la asociación entre obesidad y STC, en un diseño de casos y controles, a partir de una población diana se eligen dos grupos: los que tienen la enfermedad (casos) y los que no la tienen (controles), se mira retrospectivamente hacia atrás y se analiza el factor bajo estudio (la obesidad) o variable independiente en ambos grupos⁴⁰. 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 (<10%)^{40, 41}. A diferencia, en los estudios cohortes históricos los pacientes entran en el estudio en base a la variable bajo estudio (obesi-

Estudio longitudinal prospectivo o Cohortes Clásico**Estudio Cohortes Histórico****Estudio Cohortes Bidireccional****Figura 3. Diseño clínico.**

dad) o variable independiente denominando ese grupo como *expuestos* y se realiza un seguimiento comparando con los *no expuestos* (no obesos), observando cuantos desarrollan la enfermedad en cuestión (STC) y pudiendo analizar el riego relativo (RR).

Estudios de cohortes o longitudinales. Los estudios cohortes se clasifican en tres tipos: cohortes clásico o prospectivos; cohortes históricos o retrospectivos; y cohortes bi-direccionales, que tienen una parte del diseño que es retrospectivo y otra parte prospectiva. 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 un enfermedad (menos extendido en cirugía de la mano), o bien para el análisis de resultados de un tratamiento (outcomes), o también para analizar la validez de construcción de un instrumento de medición.⁴² Independientemente de su carácter (prospectivo, retrospectivo, o bidireccional) siempre los pacientes entran en el diseño clínico en base al factor bajo estudio o variable independiente, a diferencia de los estudios de casos-controles como veímos antes. En los casos de 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 el investigador puede calcular la mejoría tras tratamiento, teniendo en cuenta la medida antes y después de la cirugía, calculándose la magnitud de dicha mejoría clínica en base al efecto del tratamiento que se mide mediante el tamaño del efecto (ES = effect size) o la media estandarizada de la respuesta (SRM = standarized response mean); estadísticos que desarrollaremos en el apartado de «instrumentos y medidas» del material y método. En orden a ver la diferencia entre los tres tipos de cohortes (**Figura 3**), supongamos que el investigador establece el objetivo de valorar la eficacia de un tratamiento, como por ejemplo, la liberación endoscópica del ligamento transverso del carpo como tratamiento para el STC. La variable independiente que define la entrada de pacientes al estudio es el tratamiento, mientras la variable de-

pendiente es el resultado o mejoría, de ahí que esta variable también se denomine variable resultado o «variable outcome». En el caso de *cohortes clásico* el diseño será prospectivo. Se define la población estudio y se considera el momento en el tiempo para realizar las mediciones, normalmente el día antes de la cirugía, y por ejemplo 3 meses después. De forma prospectiva se van incluyendo los pacientes hasta que se reúna una población estudio con un tamaño muestral previamente analizado. A continuación, el investigador mide la magnitud de la mejoría desde antes de la cirugía hasta 3 meses después. En los *cohortes históricos*, el estudio comienza hoy, el investigador define los criterios de inclusión y exclusión que define su población, y a continuación define el intervalo mínimo de seguimiento que debe reunir cada caso para ser incluido en el estudio (3 meses). A continuación, el investigador, una vez tiene su población estudio con todos aquellos pacientes que reunían los criterios de selección y que tenían medidas 3 meses tras el tratamiento, viaja hacia atrás (retrospectivo) y recolecta las medidas pre-operatorias que ya se tenían, y, al igual que en los cohortes clásicos, mide la magnitud de la mejoría clínica. En los *cohortes bi-direccionales*, el diseño comienza como los cohortes históricos. El investigador viaja hacia atrás de forma retrospectiva, recolecta las medidas pre-operatorias, pero a diferencia de los cohortes históricos, cita a las pacientes en su unidad y recolecta de forma prospectiva las mediciones post-operatorias. Este tipo cohortes es muy útil en unidades de investigación donde de forma estandarizada se aplican mediciones pre-operatorias, y ante la generación del proyecto de investigación surge la duda si todos los pacientes tienen mediciones post-operatorias. Entonces, la solución es realizar una medición prospectiva en la unidad a todos aquellos sujetos seleccionados en el estudio para asegurar que vamos a tener medidas antes y después del tratamiento. Un ejemplo poco común de este tipo de diseño lo constituye el trabajo de uno de los autores del presente artículo donde analiza la eficacia de la reconstrucción de la inestabilidad radio-cubital distal mediante un cohortes bi-direccional⁴³.

Ensayos Clínicos Controlados Aleatorizados (ECA). Denominados también estudios experimentales, pues en ellos el investigador realiza una intervención sobre la población estudio introduciendo el elemento o variable que se quiere valorar. Todos los ECA son cohortes prospectivos en los que se comparan dos tratamientos y se quiere saber cual es el mejor; o bien se quiere analizar la eficacia de un tratamiento, y para ello se compara con un control. 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). Característicamente, 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. Esto es muy importante, pues una gran mayoría de los trabajos publicados en cirugía de la mano en los que se comparan dos tratamientos no son ECA. Así, es muy típico que en un servicio de cirugía de la mano durante un período de tiempo se use una técnica específica para tratar un determinado proceso patológico. Con el tiempo sale una nueva técnica para la misma patología y se incorpora al servicio. Pasados unos años, de forma retrospectiva, se comparan los dos grupos en los que sólo se tiene una medición en el tiempo para ver cual de los dos es mejor. Este tipo de estudios publicados tan frecuentemente no constituye un diseño clínico de ECA sino de un Estudio de Corte Transverso, en los que no podemos concluir que A es mejor que B, ya que los pacientes no han tenido la misma probabilidad de recibir un tratamiento u otro; y además no han sido valorados prospectivamente. Un ejemplo clásico es el trabajo de Kalsson et al.⁴⁴ en el que comparan la liberación simple del ligamento transverso del carpo con la reconstrucción del ligamento en el tratamiento del STC y que no constituye un ECA.

Las características principales de los ECA son: 1.- Asignación al azar de los pacientes a uno u otro grupo de tratamiento, lo cual redu-

ce 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. El grado de funcionamiento de la aleatorización es directamente proporcional al tamaño muestral: cuanto mayor sea éste, es menos probable que haya diferencias apreciables en los grupos que se comparan. 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. 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).

Aleatorización. Es el proceso por el cual los pacientes incluidos en un ECA son asignados al azar a recibir uno de los tratamientos que van a ser evaluados. Para cada protocolo de ECA, habrá un método más adecuado de aleatorización; por ello, el objetivo del presente trabajo no es describir cada uno de los pasos a seguir en la realización de un esquema de aleatorización. Sin embargo, expondremos a continuación las características más destacables de cada uno de ellos.

Aleatorización simple. Este procedimiento utiliza un código para asignar a todos los pacientes a uno u otro tratamiento al inicio del estudio. El código es generado antes de iniciar el ECA, y frecuentemente asigna los tratamientos según orden de admisión de pacientes. Su mayor ventaja es su sencillez. Sus desventajas son: a) la desigualdad del tamaño de los grupos de tratamiento, sobre todo cuando el número de pacientes de la muestra es inferior a 100 (cuanto menor es la muestra mayor probabilidad de desigualdad), lo que a su vez influye en la significación estadística; b) si el ECA es multicéntrico la distribución de los grupos de tratamiento puede que no sea equilibrada en cada centro; c) la distribución de factores pronósticos,

en caso de que existan, puede no ser homogénea en ambos grupos, lo que dificulta la interpretación de los resultados.

Aleatorización restrictiva o equilibrada. Con ella se asegura que el número de pacientes en cada grupo de tratamiento sea igual al otro, lo que aumenta el poder de los test estadísticos, sobre todo si el número total de pacientes es bajo, como ocurre en cirugía. Este tipo de aleatorización puede ser: a) *balanceada global*, que presenta las mismas dificultades que la aleatorización simple y la única ventaja es que asume un número total de pacientes conocido; b) *por bloques*, cuya principal ventaja es asegurar la igualdad en el número de pacientes en cada grupo de tratamiento, lo que supone la posibilidad de realizar análisis parciales (ej. por hospitales) y/o intermedios. La longitud de los bloques ha de ser siempre múltiplo del número de grupos de tratamiento del estudio. En general se acepta una longitud del bloque igual a 8-10, puesto que si los bloques son más largos puede ser difícil mantener la igualdad, y si son demasiados cortos es posible adivinar la asignación e incurrir en un sesgo de selección. En cualquier caso, el investigador debe conocer la longitud del bloque; c) *sistemática* en la que se repite la asignación de un primer bloque de forma secuencial, y no es muy recomendada.

Aleatorización por estratificación. Consiste en la asignación de los pacientes en base a los factores pronósticos o características especiales que puedan influir en los resultados. Es aconsejable estratificar usando no más de 2 ó 3 variables. Una vez estratificados, los pacientes se asignan de forma aleatoria a uno u otro tratamiento. Esta técnica es más importante cuando el tamaño muestral es inferior a 50.

Enmascaramiento. El mero hecho de participar en un estudio modifica, en general positivamente, la conducta de los pacientes tanto en el grupo control como experimental; es el llamado efecto «Hawthorne», nombre de una planta eléctrica en Cicero, un pueblo de Illinois, donde se realizó uno de los primeros estudios de productividad empresarial de la historia (1927-1932) por el Dr. Elton Mayo, Profesor de la Harvard Business School; el conocimiento de los empleados que estaban ba-

jo estudio hizo que aumentaran su productividad. Así, aunque la aleatorización disminuye el sesgo de selección, los resultados del estudio pueden verse distorsionados cuando el paciente, o los profesionales que participan en la realización, o en la evaluación del estudio, conocen el tratamiento utilizado. El enmascaramiento puede ser: a) *a simple ciego*, son aquellos en los que los sujetos que entran en el ECA no sabe si está 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. Un inconveniente en los ECA en cirugía es que siempre el cirujano que sigue los tratamientos tras la cirugía sabe el tratamiento que ha aplicado. Un ejemplo claro de un correcto ECA lo constituye el trabajo publicado por Atroshi et al.³³ publicado en el BMJ, en el que se compara la liberación endoscópica con la abierta estándar del túnel carpiano.

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. La frecuencia de ECA en revistas científicas relacionadas con la cirugía de la mano es muy baja. Así, Amadio et al.⁷, demostraron que, en una revisión de casi 20 años del Journal of Hand Surgery Americano, de 3107 artículos sólo se habían publicado 8 ECA. Recientemente, Gummesson et al.⁸, revisaron todos los ECA publicados entre 1992 hasta 2002 en todas las principales revistas científicas en relación a MMSS (Journal of Hand Surgery Americano, Journal of Hand Surgery Británico, Journal of Bone and Joint Surgery Americano, Journal of Bone and Joint Surgery Británico, etc.), usando la escala de Jadad para cuantificar la calidad de los ensayos clínicos publicados. Concluyeron que sólo se pudieron seleccionar 92 ECA con una calidad media con puntuación de 2 puntos (rango de la escala 0 a 5). La importancia de publicar correctamente los ECA está siendo reconocida últimamente. Una publicación inadecuada disminuye la certeza de las conclusiones

y la fuerza del mensaje que se quiere transmitir con el estudio. Un simple método para mejorar la publicación de los ECA ha sido la incorporación de las normas del CONSORT.^{45,46} En la mitad de los noventa, dos iniciativas independientes para mejorar la calidad de los ECA publicados, llevaron a la publicación de las normas del CONSORT, que fueron desarrolladas por un grupo internacional de investigadores clínicos expertos en ECA, estadísticos, epidemiólogos y editores de revistas biomédicas. El CONSORT está avalado por un creciente número de revistas científicas médicas y grupos editoriales, incluyendo: International Committee of Medical Journal Editors (ICMJE), The Vancouver Group⁴⁷, The Council of Science Editors (CSE), and the World Association of Medical Editors (WAME). Las normas del CONSORT se publican en diferentes lenguas como Holandés, Alemán, Inglés, Japonés, Francés y Español. Las normas del CONSORT⁴⁸ consisten en una lista y un diagrama de flujo para revisar 22 puntos fundamentales a la hora de enviar un manuscrito sobre ECA a una revista científica. La importancia de estas normas adoptadas por los principales editores de revista ha llevado a mejorar la información metodológica de los ECA; así en una valoración de 71 ECA publicados en tres revistas en 1994, la asignación aleatorizada y el enmascaramiento no fueron publicados claramente en el 61% de los ECA. Cuatro años más tarde, después de que estas tres revistas adoptaran las normas CONSORT para los autores, bajó espectacularmente el porcentaje de falta de información sobre la aleatorización y el enmascaramiento a un nivel del 39%. La información relacionada con los aspectos más importantes de los ECA se incluirán en los manuscritos enviados para publicación si se adoptan las normas CONSORT. La lista con los 22 puntos está disponible en internet (<http://www.consort-statement.org/Downloads/Checklist.doc>).

Meta-análisis (MA) y Revisión Sistemática (RS). Constituyen los estudios de mayor evidencia científica en investigación clínica. Los términos MA y RS son herramientas fundamentales de la Medicina Basada en la Evidencia para los profesionales médicos. Comúnmente

los lectores de revistas científicas en cirugía de la mano no están familiarizados con estos términos, existiendo a menudo confusión entre ambos. El número de trabajos científicos publicados sobre MA y RS en la literatura médica se ha incrementado rápidamente en los últimos 20 años. Sin embargo, los méritos de un procedimiento misteriosos como es el MA siguen siendo tema de debate dentro de la comunidad médica⁴⁹⁻⁵¹. Una definición muy útil fue publicada por Huque⁵² en 1998: «El MA es un análisis estadístico que combina o integra los resultados de varios ECA considerados por el analista como combinables». Sin embargo, la terminología continúa siendo discutida, y expresiones de difícil traducción como «revisión general (overview)», «análisis combinado (pooling)» y «síntesis cuantitativa (quantitative synthesis)» se siguen usando para definir estudios basados en este diseño clínico. Nosotros creemos, que el término MA debe ser reservado para describir la integración estadística de estudios separados, mientras que el término RS debería ser más apropiado para definir cualquier revisión de un conjunto de datos que usa claramente un método y criterios definidos. Así, una RS puede incluir MA; apreciaciones o revisiones de un solo ECA, y otros recursos de evidencia⁵³. 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 tanto, prevenir demoras innecesarias en la introducción de tratamientos eficaces en la práctica clínica; c) permite evaluar la hipótesis formuladas 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 se puede calcular el tamaño muestral necesario en estudios de forma precisa⁵⁴. Muy pocos MA han sido publicados en revistas científicas relacionado con la cirugía de la mano⁵⁵⁻⁵⁹, y alguno de ellos no son realmente MA^{55,59}, como el trabajo publicado por Larsen et al.⁵⁵ en el que, según los autores, intentan realizar un MA en

la literatura sobre la falta de consolidación de las artrodesis segmentarias de muñeca. Este estudio se basó puramente en trabajos publicados con un diseño clínico descriptivo, y los MA deben estar basados fundamentalmente en ECA, aunque recientemente se está extendiendo la realización de MA de estudios observacionales⁶⁰. Un claro ejemplo de un correcto meta-análisis lo constituyen los trabajos de Margaliot et al.⁵⁸, en donde comparan los resultados de la fijación externa con la fijación interna en el tratamiento de las fracturas inestables distal de radio; o el recientemente publicado por Sheina et al.⁵⁶, donde comparan tres tratamientos clásicos para la neuropatía compresiva del nervio cubital en el codo. El MA 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 si, con el fin de calcular una estimación global del efecto del tratamiento. Muy pocos investigadores se dedican a este tipo de estudios por los altos requerimientos metodológicos, por lo que recomendamos al lector como introducción al MA los trabajos de Chung et al.⁶¹ y de Petti⁶²,

A pesar de que no es frecuente la publicación de MA en revistas científicas de cirugía de la mano, existen instituciones superiores que continuamente están realizando RS y MA de la evidencia científica publicada en revistas. Una de las más importantes es La Colaboración Cochrane» (<http://www.cochrane.org/>), organización internacional sin ánimo de lucro, que prepara y mantiene RS de alta calidad para dar información certera sobre el efecto de las intervenciones sanitarias. La Librería Cochrane (<http://www.thecochranelibrary.com/>), principal producto de la Colaboración Cochrane, ha llegado a ser el mayor compendio de RS; y contiene revisiones Cochrane realizadas por los grupos de revisores Cochrane, otras RS que han sido publicadas en la literatura médica, y, finalmente, una enorme base de datos sobre ECA. Cochrane aplica una metodología científica muy rígida al proceso de síntesis de la evidencia a partir de los ECA. Recientemente, Schädel-Höpner et al.⁹ han revisado todos los MA y RS incluidos en la base de datos Cochrane relacio-

nados con la cirugía de la mano, concluyendo que de 470 RS, 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 tendones flexores, 1 sobre fracturas del cuello del 5º metacarpiano, 1 sobre dedo en martillo; y finalmente 1 sobre rizartrosis del pulgar). Las 14 RS relacionadas con la cirugía de la mano estaban comprometidas a nivel de la metodología y heterogeneidad clínica de los estudios originales incluidos en cada revisión. Así, son muchos los autores^{8,61-63} que demandan una mejoría en la calidad de los ECA en cirugía de la mano, pues sin ECA de calidad no puede haber buenas RS y MA.

C. Instrumentos y medidas

Durante muchos años de investigación clínica en cirugía de la mano, se han valorado los resultados en base a datos puramente objetivos de la exploración y examen clínico realizados por el médico o cirujano (en nuestro caso). Una afección localizada en un determinado territorio del aparato locomotor, como la mano, se puede manifestar clínicamente con pérdida de: rango de movilidad, fuerza, sensibilidad, etc... Precisamente estos parámetros objetivos eran los utilizados hasta ahora por el cirujano de mano a la hora de valorar la eficacia de los tratamientos empleados, siendo las medidas más extendidas en trabajos publicados sobre resultados en la cirugía del carpo. Pero, ¿por qué las medidas hechas por los cirujanos son las que realmente se deben emplear?⁶⁴. Toda medida está sujeta a variabilidad, hasta la realizada por la máquina más perfecta.

Por otro lado, las medidas objetivas realizadas por el propio médico 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, que pueden verse afectadas secundariamente, y que no podemos medir con los parámetros objetivos del examen clínico. Estas dimensiones de la salud relacionadas con la calidad de vida son los parámetros que preocupan al paciente, los que ne-

cesitamos mejorar con nuestros tratamientos, y los que debemos utilizar a la hora de valorar la eficacia de dichos tratamientos. 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»⁶⁵.

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 mayoría de nuestras actuaciones médicas suelen estar basadas en la experiencia, pero sólo la experiencia no implica la realización de una buena práctica médica. El examen de la evidencia procedente de la investigación clínica suplanta a la intuición y a la experiencia clínica no sistemática, y por tanto, sólo la evidencia científica implica buena praxis médica (Evidence - Based Medicine Working Group)⁶⁶.

El objetivo último del «Movimiento Outcomes» es conocer si realmente lo que hacemos en nuestra práctica médica diaria es realmente eficaz, basándonos en la evidencia científica. Uno de los más importantes problemas para llegar a este objetivo aparece a la hora de determinar qué parámetros y medidas deben ser empleados para valorar una intervención sanitaria como «eficaz». Así, resultados de un determinado tratamiento pueden ser clasificados como excelentes por el médico, pero no así por el paciente, consumidor del servicio sanitario, o como un resultado no tan bueno para el pagador de la sanidad debido, por ejemplo, a los altos costes. De esta manera, se ha introducido el uso de Instrumentos de Medición de la Calidad de Vida, basados en la opinión del paciente para la valoración de resultados en medicina. Resulta innovador el hecho de que no sólo podemos medir resultados basados en la opinión del paciente de una manera científica, sino que además, trabajos recientes^{42, 67-69}, han demostrado, que estas medidas científicas de la opinión del paciente, son más sensibles para detectar los cambios clínicos tras tratamiento que las medi-

das de la exploración física realizadas por el propio cirujano de mano a la hora de medir resultados.

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.

Tipos de Instrumentos de Salud

La mayor innovación en la valoración de resultados en las ciencias de la salud ha sido, como hemos visto anteriormente, la adopción de medidas de la salud en relación a la calidad de vida y la captura de la perspectiva del propio paciente (consumidor del servicio sanitario) sobre su salud. Este movimiento se ha extendido a casi todas las áreas de la medicina incluso a la microcirugía y a la cirugía de la mano. Cuestionarios de salud, basados en la opinión del paciente, han sido introducidos para valorar resultados en la transferencias microvasculares de dedo de pie a mano para reconstruir el pulgar⁷¹, en la cirugía abierta del túnel carpiano^{42, 67, 71}, en fracturas distales de radio⁶⁸, en la cirugía endoscópica del túnel carpiano^{33, 72}, en la cirugía de la recurrencia de síndrome del túnel carpiano⁷³, en la comparación de resultados de la cirugía del túnel carpiano en trabajadores demandantes de compensación económica con trabajadores no demandantes de compensación⁷⁴, etc. Estos cuestionarios se pueden clasificar en dos grandes grupos en base al tipo de información recolectada: genéricos y específicos.

Los instrumentos genéricos, como el conocido SF-36⁷⁵, son cuestionarios que tratan de medir todas las dimensiones importantes de la salud en relación a la calidad de vida, y pueden ser usados virtualmente 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; sin embargo, nos permite conocer el incremento general en la calidad de vida comparado con otros tipos de tratamientos 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 nuestro⁷⁶, objetivamos que, en Canarias, el detrimiento 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 en comparación a lo que se destina en una afección tan localizada en la mano como el síndrome del túnel carpiano.

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 STC) (Instrumento de Levine) (*Brigham and Women's Hospital. Boston Carpal Tunnel Instrument*)^{71,77} como representativo de instrumentos específicos de una enfermedad o condición patológica, y que mide gravedad de síntomas, estatus funcional o discapacidad en relación con la patología. Y el DASH (Disability of Arm, Shoulder and Hand), (Institute for Work & Health, AAOS, ASSH 1996)^{71,78} que es un ins-

trumento específico de una área de función, y nos sirve para medir la gravedad de los síntomas y el estado funcional de los miembros superiores (MMSS).

Instrumento 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. 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. Probablemente, es el instrumento de medición más extendido hoy en día en el campo de la cirugía de los MMSS, estando disponible en 27 lenguas (<http://www.dash.iwh.on.ca/translate.htm>). Está constituido por un cuestionario de 30 ítems. 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 DASH no cubre todos los aspectos de la salud, y se recomienda que en estudios de outcomes sea suplementado con un instrumento genérico como el SF-36, SF-12, o el Sickness Impact Profile.

El instrumento fue diseñado para ser sensible en la detección de discapacidad ocasionada por patologías que afecten cualquier parte de las extremidades superiores. Esto significó un gran avance a la hora de eliminar cuestionarios separados para medir resultados en la mano, el hombro o el codo. Se ha demostrado que el DASH es útil para describir individuos con enfermedades o patologías de los miembros superiores, 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 DASH se puntuá en dos componentes: Primero las preguntas de función/síntomas (30 ítems, puntuados 1 a 5); y segundo los módulos opcionales deportes/músicos o trabajo (4 ítems puntuados de 1 a 5).

Puntuación Función/Síntomas. Las respuestas a los primeros 30 ítems del DASH son sumados para dar «la puntuación cruda». La máxima puntuación que se puede obtener es 150, la mínima es 30. La puntuación cruda es entonces transformada a una escala de 0 a 100, donde 0 refleja la ausencia de discapacidad (buena función), y 100 muestra la mayor discapacidad (mala función).

Puntuación módulos opcionales Deporte/Música o Trabajo. Cada módulo opcional está constituido por 4 ítems que pueden ser, o no, contestados por los sujetos diana, debido a la naturaleza de las preguntas, pues están en relación con actividades específicas. 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. La puntuación cruda también es transformada a una escala de 0 a 100, en la que puntuación más baja implica discapacidad mínima, y puntuaciones altas reflejan mayor discapacidad.

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% 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).

Versión Española del Instrumento DASH. El instrumento DASH había sido desarrollado en un país con diferente lengua y cultura; y debíamos tener en cuenta las influencias socio culturales de la traducción. De esta manera, el cuestionario debía ser adaptado al castellano antes de su uso, con el fin de obtener una versión equivalente conceptualmente a la original americana. Una vez adaptado el cuestionario, se debían analizar sus propiedades psicométricas como confiabilidad y validez.

Adaptación cultural y lingüística de un Instrumento. Un instrumento para medir salud debe ser entendido como cualquier otro instrumento de medición, como por ejemplo una báscula; en este caso, tenemos un instrumento para medir la dimensión física «peso». Igualmente ocurre con los instrumentos de salud con la diferencia de que en este caso en particular vamos a medir una dimensión de la salud como la funcionalidad física, salud mental, función social, etc. Estos instrumentos de salud son simples cuestionarios desarrollados en países como Reino Unido, Canadá o EEUU. Debido a ello, el resto de los investigadores y cirujanos de mano, de habla no anglosajona, nos enfrentábamos a uno de los más importantes errores sistemáticos a la hora de usar estos instrumentos de medición, ya que solían ser traducidos por el propio investigador o cirujano para ser aplicados de inmediato en una población estudio, sin tener en cuenta los sesgos socio culturales de la traducción. Para evitar ese 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. Para entender este primer paso metodológico se expone un ejemplo clásico: imagínese que viajan a Nueva York y compran la más sofisticada báscula para medir peso

y la traen a España. No podría ser usada directamente porque la dimensión física del peso es medida en ciertos países anglosajones mediante la unidad libras y en España es kilogramos. En consecuencia sería necesario adaptar el sistema de medición de la báscula al nuevo sistema de tal forma que, cada vez que quisiéramos medir peso, supiéramos que dicha medición en Kg. es equivalente a la realizada en EEUU en libras. Existen diferentes métodos de adaptación socio cultural y lingüístico al respecto. El más utilizado es el de la traducción-retrotraducción⁷⁹, 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. La adaptación al castellano del instrumento original DASH siguió un protocolo similar al utilizado en el proyecto de la IQOLA desarrollado para obtener las diferentes versiones del instrumento genérico SF-36. El método estaba basado en la traducción-retrotraducción. Los cuestionarios originales fueron traducidos al castellano por dos traductores bilingües, con experiencia clínica, cuya lengua original era el castellano. Los traductores valoraron la dificultad para obtener una expresión en castellano conceptualmente equivalente a la original americana. Las dos nuevas versiones en castellano fueron discutidas simultáneamente entre traductores y uno de los investigadores responsables del proyecto. Otros dos traductores bilingües valoraron la equivalencia conceptual, la claridad de cada frase y cada de una de las opciones de respuesta en estas dos primeras versiones en castellano. A continuación, una segunda reunión entre traductores e investigadores fue llevada a cabo para obtener una primera versión en castellano de los cuestionarios (versión 1.0) tras llegar a un consenso. 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 ver-

siones 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. La valoración cuantitativa de la adaptación demostró que la variabilidad observada en la puntuación asignada por los traductores para la dificultad en la traducción de los ítems del instrumento fue mucho mayor comparada con la puntuación media asignada a la equivalencia conceptual.⁷¹ Finalmente, antes de que fuera presentada la definitiva versión 2.0 en castellano, se procedió a realizar el análisis de la confiabilidad (consistencia interna y reproductibilidad), validez y respuesta.

Confiabilidad de un Instrumento de medición. Una vez se tenga la versión española de un instrumento se tiene que analizar las propiedades de medición de dicho instrumento, antes de ser aplicado en una población estudio. La confiabilidad de un instrumento tiene dos conceptos: la consistencia interna y la reproductibilidad. Se entiende como consistencia interna la habilidad que tienen los ítems de una escala de un instrumento para medir un concepto común a todos ellos. Para ello se utiliza el coeficiente alfa de Cronbach, debiéndose obtener valores $>0,7$ para una buena consistencia interna (Advisory Scientific Committee of Medical Outcome Trust).⁸⁰

Se define como reproductibilidad a la estabilidad de un instrumento a través del tiempo. Por ejemplo una báscula fiable y con buena reproductibilidad sería aquella que tras dar nuestro peso, es capaz de repetir y dar la misma medición momentos después. La reproductibilidad se analiza mediante el test-retest de reproductibilidad, mediante dos mediciones separadas en un tiempo en la que se espera que la enfermedad no haya cambiado⁷⁹. Cuando en vez de usar instrumentos de salud, vamos a utilizar medidas de la exploración clínica (ej. fuerza, rango

de movilidad, etc), 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*». En toda medición para valorar resultados, debe ser estudiada su fiabilidad.

Validez de un instrumento de medición. Se define como la capacidad que presenta un instrumento para medir el concepto que propone medir. 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. Métodos comunes de acumular evidencia sobre «validez de contenido» incluyen el uso de leyes y juicios de paneles de expertos que establecen la claridad, comprensión y redundancia de ítems y escalas de un instrumento de medición. b) *Validez de criterio*: evidencia de que muestren una relación entre el instrumento nuevo que se está analizando y una medida de criterio establecida. Debido a que en el área de la salud no existen «medidas criterios» válidas y establecidas, la validez de criterio raramente es estudiada. Así recientemente Alexander et al.⁸¹, intentando analizar la validez de un nuevo instrumento de medición para medir las modernas actividades de la mano en relación a las nuevas tecnologías (ej. el uso del móvil, etc.), establecieron un estudio de validez de criterio correlacionando las mediciones de este nuevo instrumento con el DASH, usando éste último como medida de criterio de la función de la mano; cuando los propios autores que desarrollaron el DASH expresaron en sus publicaciones originales que les fue imposible analizar la validez de criterio del DASH porque no existían medidas-criterio en la salud.⁸¹ Este análisis de validez (realizado por estos autores de Harvard) constituye un típico error conceptual, ya que la validez analizada fue una validez de construcción. c) *Validez de construcción*. Es una de las más utilizadas por 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^{22, 37, 42}. Si-

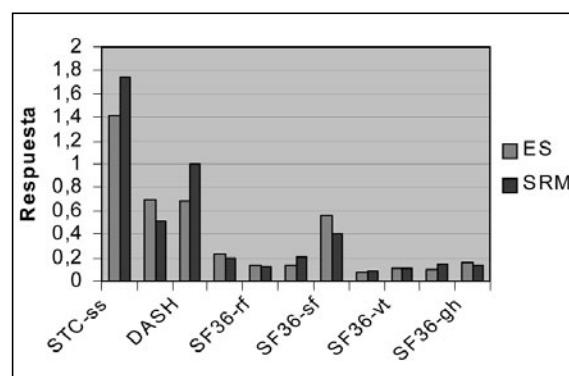


Figura 4. Validez de Construcción del Instrumento DASH en base al análisis relativo de la respuesta comparada con el SF- 36 y el instrumento específico STC (ES= tamaño del efecto) (SRM= media estandarizada de la respuesta)

guiendo con el ejemplo de la báscula, para conocer si las mediciones con nuestra nueva báscula (instrumento de medición) presentan una validez de construcción, sólo tendríamos que relacionarlas con otras medidas como las del volumen corporal. Dicho de otra manera, si la báscula detecta un incremento en el peso que mide, cabría esperar una relación lógica por ejemplo con un incremento del volumen corporal, medido con otro instrumento de probada validez. La gran mayoría de los trabajos científicos encaminados a valorar la validez de construcción se basan en el análisis de la correlación matemática que existiría entre la medición con el nuevo instrumento con respecto a la medición realizada en la misma población estudio pero con otro instrumento de probada validez. Sin embargo, nosotros establecimos 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 (específico de MMSS) 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 será menor a la que pudiera presentar un instrumento específico de la enfermedad como el cuestionario de Levine⁷⁷ específico de STC (**Figura 4**). Al igual que veíamos antes con la fiabilidad, las medidas de la validez son muy escasas en trabajos publicados en cirugía de la mano en los que usan me-

didas de la exploración física. Es muy frecuente encontrar trabajos en los que los resultados son clasificados (excelentes, buenos, regulares y malos) en base a sistemas o clasificaciones de las que no se ha demostrado su validez. Así, cuando clasifico un resultado como excelente, debo tener pruebas sobre la validez de que esa asignación de «resultado excelente» es realmente excelente.

Respuesta de un instrumento de medición. Se conoce como respuesta a la sensibilidad de una medida o instrumento de medición para detectar cambios de importancia clínica. 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. Del 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 cuales son las medidas de elección a emplear para esa población específica y para el tiempo estimado de seguimiento o de 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 = 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 = 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 tra-

tamiento y su significado está en que nos dice que diferencias estadísticas significativas tienen, o no, importancia clínica. Así, recientemente hemos demostrado que las medidas del DASH, y de otros instrumentos 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 STC⁴².

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. Existe una versión corta del DASH denominada QuickDASH, con 11 ítems, de la que existen muy pocos trabajos publicados²², pudiéndose usar en vez del DASH estándar con la misma fiabilidad, validez y respuesta (<http://www.dash.iwh.on.ca/translate.htm>).

Aplicación del DASH en la cirugía de la mano. La versatilidad del instrumento DASH hace posible su uso en innumerables patologías del miembro superior. Sus primeros usos se centraron en problemas que afectaban las grandes articulaciones proximales como hombro y codo, de hecho los primeros estudios sobre confiabilidad y validez se realizaron en estas poblaciones diana⁸³. A nivel de la mano, se ha empleado para valorar resultados en el síndrome del túnel carpiano^{33, 37, 42, 67, 68, 71}, en fracturas distales de radio⁸⁴, en el uso de colgajos libres para reconstrucción de lesiones graves de los nervios digitales⁸⁵, en el tratamiento de la enfermedad de Kienböck^{86, 87}, para valorar la eficacia de específicas artrodesis segmentarias del carpo, como la fusión de las cuatro esquinas en comparación con la carpectomía proximal⁸⁸, etc. Un aspecto olvidado en el uso del DASH a la hora de valorar resultados, es que este instrumento fue diseñado para valorar resultados en diseños clínicos tipo cohortes con medidas antes y después del tratamiento⁸². Lo importante es valorar la magnitud del cambio clínico tras tratamiento, y muchos de los tra-

Tabla IV - PRUEBAS ESTADÍSTICAS PARA LA COMPARACIÓN DE LA MEDIA EN BASE A LA EXISTENCIA DE UNA EXPLORACIÓN DE LAS MUESTRAS Y SU DISTRIBUCIÓN NORMAL O NO

		2 muestras		> 2 muestras		Test a posteriori (Scheffé, Tukey, LSD, SNK, Duncan)	
Paramétrica	Ligadas	Independientes	Independientes	1 fuente de variación	2 fuentes de variación		
	t-Student (datos ligados) o test de las diferencias de "cero" (t)	t-Student (t)		ANOVA I (F)	ANOVA II (F)		
No paramétrica	Test de comparación de medianas		Independientes	Test de Kruskal-Wallis (KW / H)	ANOVAS SOFISTICADAS		
	Test de Wilcoxon (T)	Test de Mann-Whitney (Z, W)		Test de Friedman (S)			

Test de comparación de tendencias de Kolmogorov-Smirnov

bajos publicados en los que se ha usado este instrumento⁸⁵⁻⁸⁷, basan sus resultados en una sola medición, lo cual constituye un error metodológico y los resultados publicados carecen de valor. El presentar como resultado que el instrumento DASH mostró tras tratamiento una puntuación media de 5 carece de valor científico, y mucho menos cuando dicha puntuación es clasificada en discapacidad leve, moderada o grave, ya que no existen evidencia que demuestren la validez de dicha gradación con la puntuación del DASH.

D. Análisis Estadístico.

Los métodos estadísticos son una herramienta fundamental en la investigación clínica, tanto en las etapas de diseño como en los proce-

sos de selección de muestras y análisis de los resultados. Una vez hemos establecido nuestra pregunta de investigación y el propósito del trabajo, definida nuestra población estudio, establecido el diseño clínico, y elegidas las medidas e instrumentos de medición a emplear, entramos en el análisis de los datos obtenidos. En este punto debemos comenzar con definir todas las variables de las que hemos recogido datos y clasificarlas de una forma matemática. Así, las variables numéricas o cuantitativas son aquellas que pueden cuantificarse. Se pueden dividir en: a) discretas, si no admiten tomar todos los valores intermedios en un rango, tomando con frecuencia valores enteros (número de hijos, o número de partos); y b) continuas, si admiten tomar cualquier valor dentro de un rango determinado (edad, peso, índice de

masa corporal, puntuación del DASH, fuerza, rango de movimiento, etc). Las variables cualitativas (o categóricas) representan una cualidad o atributo que clasifica a cada individuo en una o varias categorías. Se dividen en: a) ordinales, si las categorías son susceptibles de ser ordenadas de modo lógico (clase social: baja, media, alta); y b) nominales, si las categorías no siguen ningún orden natural (grupo sanguíneo: A, B, AB, O). Las variables cualitativas que sólo aceptan dos posibilidades se denominan dicotómicas (género: hombre/mujer). Del investigador depende como se codifica las variables a estudiar, así, una variable continua como la edad se puede convertir en una variable categórica separando en intervalos la variable edad (ej.: <30 años, 30-45, 46-60, >60 años). Es importante tener una hoja de codificación que siempre nos explique como se ha descargado los datos, pues el ordenador y los programas actuales de análisis estadísticos sólo entienden de números; así una variable dicotómica (hombre/mujer) se debe codificar a valores numéricos (hombre = 1; mujer = 2). En caso contrario, ni siquiera el estadístico podrá realizar el análisis.

Una vez descargados los datos de todas las variables, debemos determinar aquellas variables independientes (edad, género, tiempo transcurrido en instaurar el tratamiento, tiempo de seguimiento, etc.); y diferenciarlas de las variables resultado, llamada también variable dependiente o variable «outcomes». Por ejemplo, supóngase que el investigador quiere establecer si existe relación causal entre la edad, género y la actividad manual con el desarrollo de STC, en este análisis multivariante, la variable dependiente será tener o no STC, basados en los criterios estimados por el investigador, mientras que las variables independientes serían el género, la edad, y el grado de actividad manual. En otro ejemplo, el investigador realiza un ECA y el objetivo es comparar la LETC con la LATC, así debe establecer cuales son las variables resultados (principales y secundarias) a la hora de realizar el análisis estadístico; ej. la gravedad de síntomas (medida con el instrumento STC); la discapacidad en MMSS (medida con el DASH), etc., todas ellas constituyen variables dependientes o «outcomes».

El análisis estadístico debe comenzar con la exploración de la/s muestra/s, con análisis descriptivo de las variables, comenzando con las variables básicas, que definen la población estudio como edad, género, mano afecta, etc., y dando datos sobre la tendencia central del análisis descriptivo de cada variable (media, mediana, moda), y a continuación aportar datos sobre la variabilidad de los datos en cada variables con medidas de la dispersión (varianza, DS, error estándar sobre la media, rango con mínimo y máximo, etc). Esto último es de suma importancia, pues es frecuente observar trabajos de investigación clínica donde aportan como medida de un resultado la «media» pero sin ninguna información sobre la dispersión de los datos (ej. DS o error estándar sobre la media), ya que para el lector con experiencia en investigación clínica le puede dar información sobre si los test estadísticos empleados son adecuados, y por ende si los resultados del trabajo pueden ser aceptados como evidencia científica. El ejemplo más claro: supóngase que un investigador se ha propuesto valorar la eficacia de un tratamiento mediante un cohortes clásico con mediciones antes y después del tratamiento. Establece como análisis estadístico para valorar resultados la comparación de la media de la variable continua «gravedad de los síntomas» mediante un prueba de t-Student para muestras apareadas (medición antes y medición después del tratamiento) y observamos que la DS es mucho mayor que la media en cada medición, ello nos apuntaría la posibilidad de que las muestras estudiadas no siguen una distribución normal y en ese caso se debería haber usado un test de comparación de medias no paramétrico para muestras apareadas. Por ello, el paso fundamental antes de realizar los test estadísticos propios del objetivo del estudio es explorar la/s muestra/s, en orden a conocer si sigue una distribución normal, o, lo que es lo mismo, que el 95% de los datos se encuentran agrupados en dos DS alrededor de la media (Prueba de Kolmogorov-Smirnov, Prueba de Shapiro-Wilk's), y analizar la homocedasticidad de las varianzas si vamos a comparar dos medias (Test de Levine) para saber si tenemos que usar un test paramétrico o no-paramétrico. En la **Tabla IV** se aporta informa-

ción básica sobre los test paramétricos y no paramétricos a usar.

RESULTADOS

Esta parte de un artículo de investigación clínica es en la que el/los autor/es exponen de forma directa lo descubierto con su investigación. Es importante no hacer comentarios o interpretación de los resultados obtenidos, pues ello debe ser incluido en la siguiente sección denominada «Discusión». Así, la sección de los resultados sólo debiera incluir datos, no referencias ni citas bibliográficas a otros trabajos. Tablas y gráficos deben ser usadas si con ello hacemos más claro, o más fácil de entender, los datos. Es recomendable que los autores comiencen la presentación de los resultados con una tabla sobre las características de la muestra con distribución de frecuencias, medias, medianas y DS. En estudios observacionales y experimentales (ECA) aportar información sobre las «perdidas» en el seguimiento, y si estos difieren de los sujetos sometidos a observación en el estudio. Cuando se usan tablas para presentar los resultados deberían tenerse en cuenta los siguientes puntos: 1.- presentar las figuras y tablas más claras y comprensibles; 2.- las tablas y figuras deben tener títulos claros y estar suficientemente explicadas de manera de que los lectores puedan comprenderlas e interpretarlas sin necesidad de acudir al texto general de manuscrito, sin embargo, eso no implica que repitamos la misma descripción de los resultados en el texto y en las tablas o gráficos; 3.- especificar las variables estudiadas y las unidades empleadas en su medición (ej. tiempo de seguimiento: años, meses, semanas); 4.- en caso de resultados estadísticamente significativos aportar el valor real de «p» (ej. en vez de $p<0.05$, presentar $p=0.02$); 5.- No ser redundante (ej. si

aportamos la media y la DS como medida de la dispersión de datos no es necesario aportar también intervalos de confianza al 95%, el máximo y el mínimo, etc.).

DISCUSIÓN

Es la sección de un trabajo de investigación donde los autores interpretan los resultados observados y presentados en la sección anterior de «resultados». La discusión debe empezar con un párrafo que sintetice o resuma los resultados obtenidos. A continuación debe comparar sus resultados con la evidencia publicada, pero no en términos generales, sino de manera precisa. Aquí es el momento de incluir todas las referencias importantes con los datos publicados por otros autores. Una vez comparados nuestros resultados, es la hora de interpretar los mismos con el fin de contestar la pregunta de investigación que formulamos en la sección inicial de «introducción». Antes del final de la discusión, los autores deben incluir al menos un párrafo donde se expliquen las limitaciones que presenta su trabajo de investigación, y las futuras investigaciones que se proponen con el fin de seguir avanzando en el conocimiento de la respuesta a la pregunta de la investigación que fue el objetivo del trabajo.

AGRADECIMIENTOS

Los autores desean expresar su agradecimiento a la Dra. Dña. Isabel Díez de la Lastra Bosch por su colaboración en la corrección del presente trabajo. Al Prof. D. Manuel Nogales del CSIC por sus enseñanzas en el campo de la bio-estadística y por la aportación de la Tabla IV de este artículo.

BIBLIOGRAFÍA

- Kleinert HE, Kasdan ML, Romero JL. Small blood-vessel anastomosis for salvage of severely injured upper extremity. *J Bone Joint Surg Am*, 1963; 45: 788-96.
- Early PF. Population studies in Dupuytren's Contracture. *J Bone Joint Surg Br*, 1962; 44: 602-13.
- Nuno-Alegre D, Azevedo L, Ferreira N, et al. Doença de Dupuytren. Revisão de 100 doentes operados. *Rev Iberoamer Cir Mano*, 2000; 34: 49-59.
- Zancolli EA, Zancolli ER, Cagnone JC. Rizartrosis del pulgar. Tratamiento quirúrgico en estados iniciales y tardíos. *Rev Iberoamer Cir Mano*, 2000; 27: 8-18.
- Irisarri C. Aetiology of Kienböck's disease. *J Hand Surg Br*, 2004; 29: 279-85.
- Pérez de la Fuente T, Vega García C, Alvelay Laso O, et al. Tumor de células gigantes de vainas tendinosas: una serie retrospectiva de 107 casos. *Rev Iberoamer Cir Mano*, 2003; 30: 29-33.
- Amadio PC, Higgs P, Keith M. Prospective comparative clinical trials in The Journal of Hand Surgery American, *J Hand Surg Am*, 1996; 21: 925-9.
- Gummesson C, Atroshi I, Ekdahl C. The quality of reporting and outcome measures in randomized clinical trials related to upper-extremity disorders. *J Hand Surg Am*, 2004; 29: 727-34.
- 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.
- Guallar E, Conde J, de la Cal MA, et al. Guía para la evaluación de proyectos de investigación en Ciencias de la Salud. Fondo de Investigación Sanitaria, Instituto Carlos III, Madrid: Ministerio de Sanidad y Consumo, 1995.
- Hulley SB, Cummings SR. Diseño de la investigación clínica. Un enfoque epidemiológico. Barcelona: Doyma, 1993.
- Roca J. Cómo y para qué hacer un protocolo. *Med Clin (Barc)*, 1986; 87: 816-9.
- McCabe SJ, Uebel AL, Pihur V, et al. Epidemiologic associations of carpal tunnel syndrome and sleep position: Is there a case for causation? *Hand*, 2007; 2: 127-34.
- 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.
- Granados A. Evaluación de tecnologías y mejora de la eficiencia del sistema sanitario. En: Evaluación del Impacto de las Intervenciones Sanitarias. Asociación Española de Evaluación de Tecnologías Sanitarias, Madrid: Artur SA, 1998: 69-71.
- Chung KC, Walters MR, Greenfield ML, et al. Endoscopic versus open carpal tunnel release: a cost-effectiveness analysis. *Plast Reconstr Surg*, 1998; 102: 1089-99.
- Silman AJ. Epidemiological studies: a practical guide. Cambridge: University Press, 1995.
- Gunnarsson LG, Amilon A, Hellstrand P, et al. The diagnosis of carpal tunnel syndrome. Sensitivity and specificity of some clinical and electrophysiological tests. *J Hand Surg Br*, 1997; 22: 34-37.
- Atroshi I, Johnsson R. Evaluation of portable nerve conduction testing in the diagnosis of carpal tunnel syndrome. *J Hand Surg Am*, 1996; 21: 651-4.
- Atroshi I, Gummesson C, Johnsson R, et al. Severe carpal tunnel syndrome potentially needing surgical treatment in a general population. *J Hand Surg Am*, 2003; 28: 639-44.
- Ozyürekçü T, McCabe SJ, Goldsmith LJ, et al. The minimal clinically important difference of the Carpal Tunnel Syndrome Symptom Severity Scale. *J Hand Surg Am*, 2006; 31: 733-8.
- Gummesson C, Ward MM, Atroshi I. The shortened disabilities of the arm, shoulder and hand questionnaire (QuickDASH): validity and reliability based on responses within the full-length DASH. *BMC Musculoskelet Disord*, 2006; 7: 44.
- Pliefke J, Stengel D, Rademacher G, et al. Diagnostic accuracy of plain radiographs and cineradiography in diagnosing traumatic scapholunate dissociation. *Skeletal Radiol*, 2008; 37: 139-45.
- Atroshi I, Gummesson C, Johnsson R, et al. Prevalence of carpal tunnel syndrome in a general population. *JAMA*, 1999; 282: 153-8.
- Rosales RS. Síndrome de Doble Compresión. En: García-López A, Rosales RS, López-Durán L (Eds). *Neuropatías Compresivas y de Atrapamiento*. Madrid: Momento Médico, 2007:197-215.
- World Health Organization. *IDC-10. International Classification of the Disease and Related Health Problems*. Tenth revision. Second edition. World Health Organization. Geneva, 2005.
- Jelsma J. Use of the International Classification of Functioning, Disability and Health. A literature survey. *J Rehabil Med*, 2009; 41: 1-12.
- Canadian Task Force on the periodic health examination: The periodic health examination

- tion. *CMAJ*, 1979; 121: 1193-254.
29. Sackett DL. Rules of evidence and clinical recommendations on use of antithrombotic agents. *Chest*, 1986; 89 (Suppl. 2): 2S-3S.
30. Cook DJ, Guyatt GH, Laupacis A, et al.. Clinical recommendations using levels of evidence for antithrombotic agents. *Chest*, 1995; 108 (Suppl. 4): 227S-230S.
31. Yusuf S, Cairns JA, Camm AJ, et al. Evidence-Based Cardiology. London: BMJ Publishing Group, 1998.
32. Chow JC, Hantes ME. Endoscopic carpal tunnel release: Thirteen years' experience with the Chow Technique. *J Hand Surg Am*, 2002; 27: 1011-8.
33. Atroshi I, Larsson GU, Ornstein E, et al. Outcomes of endoscopic surgery compared with open surgery for carpal tunnel syndrome among employed patients: randomised controlled trial. *BMJ*, 2006; 332: 1463-4.
34. Grimes DA, Schulz KF. Descriptive studies: what they can and cannot do. *Lancet*, 2002; 359: 145-9.
35. Sánchez Rosales S, Rodríguez Pulido F, Sánchez Rosales R. Síntomas, Discapacidad de MMSS y Calidad de Vida en el músico. *Rev Iberoamer Cir Mano*, 2005; 33: 75.
36. Atroshi I, Gummesson C, Johnsson R, et al. Median nerve latency measurement agreement between portable and conventional methods. *J Hand Surg Br*, 2000; 25: 73-7.
37. Atroshi I, Gummesson C, Andersson B, et al. The disabilities of the arm, shoulder and hand (DASH) outcome questionnaire: reliability and validity of the Swedish version evaluated in 176 patients *Acta Orthop Scand*, 2000; 71: 613-8.
38. Atroshi I, Gummesson C, Ornstein E, et al. Carpal tunnel syndrome and keyboard use at work: a population-based study. *Arthritis Rheum*, 2007; 56: 3620-5.
39. Stallings S, Kasdan M, Soergel B, et al. A case-control study of obesity as a risk factor for carpal tunnel syndrome in a population of 600 patients presenting for independent medical examination. *J Hand Surg Am*, 1997; 22: 211-5.
40. Wacholder S, McLaughlin JK, Silverman DT, et al. Selection of controls in case-control studies. I. Principles. *Am J Epidemiol*, 1992; 135:1019-28.
41. Schlesselman JJ. Case-control studies. Design, conduct, analysis. New York: Oxford University Press, 1982.
42. Rosales RS, Diez de la Lastra I, McCabe SJ, et al. Responsiveness and Construct validity of the DASH instrument for outcomes assessment in open carpal tunnel release. *J Hand Surg Eur*, 2009; 34: 72-75.
43. Tsuruta T, Rosales RS, Schecker LR. Evaluation of reconstruction of the unstable distal ulna. *Hand Surg*, 1996; 1: 167-75.
44. Kalsson HK, Hagberg L. Ligament lengthening compared with simple division of the transverse carpal ligament in the open treatment of CTS. *Scand J Plast Reconstr Surg Hand Surg*, 1997; 31: 65-9.
45. Sauerland S, Davis TR. The Consolidated Standards of Reporting Trials (CONSORT): better presentation of surgical trials in the *J Hand Surgery Br*, 2004; 29: 621-4.
46. Chung KC. Commentary: the quality of reporting and outcome measures in randomized clinical trials related to upper-extremity disorders. *J Hand Surg Am*, 2004; 29:735-7.
47. Davidoff F. News from the International Committee of Medical Journal Editors. *Ann Intern Med*, 2000; 133: 229-31.
48. Moher D, Schulz KF, Altman DG, and the CONSORT Group. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *Lancet*, 2001; 357: 1191-4.
49. Naylor CD. Meta-analysis and the meta-epidemiology of clinical research. *BMJ*, 1997; 315: 617-9.
50. Bailar JC. Editorial: The promise and problems of meta-analysis. *N Engl J Med*, 1997; 337: 559-61.
51. Editorial: Meta-analysis under scrutiny. *Lancet*, 1997; 350: 675.
52. Huque MF. Experiences with meta-analysis in NDA submissions. *Proceedings of the Biopharmaceutical Section of the American Statistical Association*, 1988; 2: 28-33.
53. Chalmers I, Altman DG. Foreword. En: Chalmers I, Altman DG (Eds). *Systematic Reviews*. London: BMJ Publishing, 1995.
54. Davey Smith G, Egger M, Phillips M, et al. *Meta-análisis*. Asociación Española de Evaluación de Tecnologías Sanitarias. Madrid: Artur SA, 1998.
55. Larsen K, Jacoby R, McCabe S. Non-union rates of the limited carpal arthrodesis: a meta-analysis of the literature. *J Hand Surg Am*, 1997; 22: 66-73.
56. Macadam SA, Gandhi R, Bezuohly M, et al. Simple decompression versus anterior subcutaneous and submuscular transposition of the ulnar nerve for cubital tunnel syndrome: A meta-analysis. *J Hand Surg Am*, 2008; 33: 1314. e1-e12.
57. Merrell GA, Barrie KA, Katz D L, et al. Results of nerve transfer techniques for restoration of shoulder and elbow function in the context of a meta-analysis of the English literature. *J Hand Surg Am*, 2001; 26: 303-14.

58. Margaliot Z, Haase SC, Kotsis SV, et al. A meta-analysis of outcomes of external fixation versus plate osteosynthesis for unstable distal radius fractures. *J Hand Surg Am*, 2005; 30: 1185e1-e17.
59. Merrell GA, Wolfe S, Slade JF. Treatment of scaphoid nonunions: Quantitative meta-analysis of the literature. *J Hand Surg Am*, 2002; 27: 685-91.
60. Egger M, Schneider M, Davey Smith G. Meta-analysis spurious precision?. Meta-analysis of observational studies. *BMJ*, 1998; 316: 140-4.
61. Chung KC, Burns PB, Kim HM. A practical guide to meta-analysis. *J Hand Surg Am*, 2006; 31: 1671-8.
62. Petti DB. Meta-analysis, decision analysis, and cost-effectiveness. Methods for quantitative synthesis in medicine. New York: Oxford University Press, 1994.
63. Gerritsen AA, de Vet HC, Scholten RP, et al. Enabling meta-analysis in systematic reviews on carpal tunnel syndrome. *J Hand Surg Am*, 2002; 27: 828-32.
64. Amadio P.C. Outcome assessment in hand surgery. *Clin Plast Surg*, 1997; 24: 191-4.
65. Epstein AM. The Outcome Movement - Will it get us where we want to go?. *N Engl J Med*, 1990; 323: 266-70.
66. Evidence Based Medicine Group. Evidence-Based Medicine. A new approach to teaching the practice of medicine. *JAMA*, 1992; 266: 240-2.
67. Amadio P, Silverstein M, Ilstrup D, et al. 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.
68. Amadio P, Silverstein M, Ilstrup D, et al. Outcome after Colles fracture: The relative responsiveness of three questionnaires and physical examination measures. *J Hand Surg Am*, 1996, 21: 781-7.
69. Katz JN, Gelberman RH, Wright EA, et al. Responsiveness of self-reported and objective measures of disease severity in carpal tunnel syndrome. *Med Care*, 1994; 32: 1127-33.
70. Chung KC, Arbor A, Wei F. An outcome study of thumb reconstruction using microvascular toe transfer. *J Hand Surg Am*, 2000; 25: 651-8.
71. Rosales RS, Bensenny-Delgado 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.
72. Atroshi I, Johnsson R, Sprinchorn A. Self-administered outcome instrument evaluated in carpal tunnel syndrome. *Acta Orthop Scand*, 1998; 69: 82-8.
73. Cobb TK, Amadio PC, Leatherwood DF, et al. Outcome of reoperation for carpal tunnel syndrome. *J Hand Surg Am*, 1996; 21: 347- 56.
74. Atroshi I, Johnsson R, Nouham R, et al. Use of outcomes instruments to compare worker's compensation and non-worker's compensation carpal tunnel syndrome. *J Hand Surg Am*, 1997; 22: 882-8.
75. 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.
76. Rosales RS. Uso de instrumentos de salud para valorar resultados en las neuropatías por compresión. En: Fundación Mapfre Medicina. eds. *Neuropatías periféricas por compresión*. Editorial Mapfre. Madrid, 2002; 1-26.
77. Levine DW, Simmons BP, Koris MJ, et al. A self-administered questionnaire for assessment of severity of symptoms and functional status in carpal tunnel syndrome. *J Bone Joint Surg Am*, 1993; 75: 1585-92.
78. Beaton DE, Katz JN, Fossel AH, et al. 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.
79. Ware JE, Gandek BL, Keller SD, et al. Evaluating instrument cross-nationally. methods from the IQOLA project. En: Spiker B (Ed). *Quality of Life and Pharmacoeconomics in Clinical Trials*. Second Edition. Philadelphia: Lippincott-Publisher, 1996: 681-92.
80. Advisor Committee of Medical Outcome Trust. *Instrument Review Criteria*. Medical Outcomes Trust. Sources pages, 1997; 6-10.
81. Alexander M., Franko OI, Makhni EC, et al. Validation of a modern activity hand survey with respect to reliability, construct and criterion validity. *J Hand Surg Eur*, 2008; 33: 653-60.
82. Solway S, Beaton DE, McConnell S, et al. The DASH outcome measure user's manual. Second edition. Toronto, Ontario: Institute for Work and Health, 2002.
83. Turchin DC, Beaton DE, Richards RR. Validity of observer-based aggregate scoring systems as descriptors of elbow pain, function, and disability. *J Bone Joint Surg Am*, 1998; 80: 154-62.
84. Mcdermid JC, Richards RS, Donner NB, et al. Responsiveness of the SF-36 and hand questionnaire, and physical im-

- pairment measurements in evaluating recovery after a distal radius fracture. *J Hand Surg Am*, 2000; 25: 330-40.
85. del Piñal F, García-Bernal FJ, Regalado J, et al. The tibial second toe vascularized neurocutaneous free flap for major digital nerve defects. *J Hand Surg Am*, 2007; 32: 209-17.
86. Tambe A, Ali F, Trail I, et al. Is radiolunate fusion a viable option in advanced Kienböck disease? *Acta Orthop Belg*, 2007; 73: 598-603.
87. Tambe AD, Trail IA, Stanley JK. Wrist fusion versus limited carpal fusion in advanced Kienböck's disease. *Int Orthop*, 2005; 29: 355-8.
88. Cohen MS, Kozin SH. Degenerative Arthritis of the wrist: proximal row carpectomy vs scaphoid excision & four corner fusion. *J Hand Surg Am*, 2001; 26: 94-104.