

# Valoración de resultados en Cirugía Ortopédica y Traumatología

E. Cáceres-Palou, F. López-Prats, M. Mesa-Ramos, J. Sánchez-Sotelo y S. Suso-Vergara  
 IMAS (Hospitales del Mar y la Esperanza). Barcelona.

**Introducción.** Existe un interés creciente en nuestra especialidad por el empleo de instrumentos de valoración del estado de salud para la evaluación de resultados. La esencia del quehacer médico consiste en determinar el grado de malestar del paciente, identificar los problemas que amenazan su bienestar (establecer un diagnóstico), recomendar acciones que ayuden a restablecer su salud (establecer una indicación) y ejecutar dichas acciones.

**Estado actual.** En este artículo revisaremos el estado actual del tema tras un profundo análisis de la bibliografía.

**Conclusiones.** En resumen, para la medición de la función específica en la columna vertebral, los cuestionarios de Roland y Morris y el índice de discapacidad de Oswestry son los más recomendables. El SF-36 es el más útil para evaluar el estado de salud global y el dolor crónico. Para el estatus laboral puede utilizarse la dimensión funcional del SF-36 o escalas más específicas como el WL-26 y finalmente el PSS para la satisfacción del paciente.

**Palabras clave:** resultados, valoración, cirugía ortopédica.

## Assessment of results in Orthopaedic

**Introduction.** There is growing interest among practitioners in our specialty in the use of instruments to assess health status for outcome evaluation. The essence of the physician's task is to evaluate the patient's discomfort, identify problems that threaten their well-being (establish a diagnosis), recommend actions that help to restore health (establish the indication), and execute these actions.

**State of the art.** We assessed the state of the art after a review of the literature.

**Conclusions.** Specific spinal column function is measured best using the Roland and Morris questionnaires, and the Oswestry disability index. SF-36 is the most useful instrument for evaluating overall health and chronic pain. Occupational status can be assessed with the SF-36 or more specific scales like WL-26. Finally, the PSS is useful for assessing patient satisfaction.

**Key words:** results, assessment, Orthopaedic Surgery.

Existe un interés creciente en nuestra especialidad por el empleo de instrumentos de valoración del estado de salud para la evaluación de resultados. La esencia del quehacer médico consiste en determinar el grado de malestar del paciente, identificar los problemas que amenazan su bienestar (establecer un diagnóstico), recomendar acciones que ayuden a restablecer su salud (establecer una indicación) y ejecutar dichas acciones. El actuar médico se debe considerar un éxito sólo si contribuye a mejorar el bienestar del paciente. Por lo tanto, la determinación del estado de salud resulta crucial en las diversas fases del quehacer médico y es inherente a la práctica médica<sup>1,2</sup>.

### Correspondencia:

E. Cáceres-Palou.  
 Servicio COT. IMAS.  
 Hospitales del Mar y la Esperanza.  
 Passeig Marítim, 25-29.  
 08003 Barcelona.  
 Correo electrónico: ecaceres@imas.imim.es

En el entorno de la corriente de la medicina basada en la evidencia, el estudio de diversos metaanálisis de la literatura ortopédica publicada han puesto de manifiesto diferencias terapéuticas importantes sobre temas tan diversos como la estenosis de canal, la artrodesis vertebral, la epicondilitis o las fracturas de la cadera. Esto explica en gran parte las variaciones de la práctica médica repetidamente observadas en muchas de las patologías que tratamos y la gran variabilidad en los resultados finales de gran parte de procesos específicos. En la literatura anglosajona se emplea el término *outcome* (resultado final o efecto) para designar el impacto que un determinado proceso o tratamiento tiene sobre el bienestar del paciente, es decir, cómo el proceso o tratamiento modifican el estado de salud o la calidad de vida relacionada con la salud de una persona. A diferencia de los parámetros clínicos la determinación del resultado final es un proceso centrado en la valoración que el propio paciente hace de su estado de salud. Por lo tanto, el resultado final se mide con instrumentos de valoración del estado de salud, no con parámetros clínicos. Hasta que los cirujanos ortopédi-

cos no realicemos un análisis metodológicamente correcto de nuestros tratamientos con instrumentos para valorar el estado de salud, careceremos de una información sólida en la que basar nuestra práctica médica.

## LA VALORACIÓN DE RESULTADOS EN CIRUGÍA ORTOPÉDICA Y TRAUMATOLOGÍA: ESTADO ACTUAL

La investigación de los resultados se diferencia de la investigación clínica más tradicional en el rigor y la variedad de la metodología utilizada y en la naturaleza de los objetivos finales elegidos. La práctica de una adecuada táctica científica para alcanzar los objetivos de cualquier tipo de estudio (experimental, observacional, etc.), es cada vez más frecuente<sup>3</sup>. La medicina basada en la evidencia es ahora parte del léxico ortopédico y muchas de nuestras publicaciones ortopédicas actuales clasifican la calidad metodológica del trabajo, y por tanto, la calidad científicotécnica del mismo.

Actualmente se continúa ofertando un tratamiento quirúrgico en base a la experiencia, pero la experiencia aislada no implica buena praxis médica; sólo la evidencia científica implica buena práctica médica. De esta manera, la investigación clínica en Cirugía Ortopédica y Traumatología (COT) se convierte en un verdadero producto de mercado al poder ofertar a los pacientes aquello que científicamente es útil para obtener el mejor resultado, mejorando las probabilidades de éxito del tratamiento, y por tanto realizar una medicina más competitiva. Este es precisamente el objetivo último del movimiento *Outcomes*. Se debe hacer aquello que realmente se ha demostrado que es eficaz para el paciente.

Para un paciente que sufre un determinado proceso con la consiguiente modificación de su calidad de vida, la eficacia del tratamiento que recibe le aporta cierta satisfacción. Esta satisfacción representa el grado en que la atención prestada, durante el proceso, satisface las expectativas del paciente para mejorar y/o recuperar su calidad de vida. Hay que tener en cuenta que no siempre guarda una relación directa con el nivel de calidad científicotécnica, puesto que depende en gran medida del éxito obtenido en el proceso de relación interpersonal. La más impecable y arriesgada de las intervenciones quirúrgicas puede no ser apreciada por un paciente que no ha establecido una buena relación con su cirujano. Por el contrario, un paciente que ha sufrido, por ejemplo, una infección en una artroplastia de la cadera puede quedar eternamente agradecido al equipo de profesionales que le ha atendido si considera que el trato recibido a nivel personal superaba sus expectativas.

Para Donabedian<sup>4</sup> «la satisfacción del paciente debe ser uno de los resultados deseados de la asistencia, incluso un elemento del propio estado de salud». Según este autor, la validez de la satisfacción del paciente como medida de calidad es incuestionable, y la información que se obtenga so-

bre esta dimensión debería ser indispensable para la planificación de la asistencia sanitaria.

Por tanto, de un sistema clásico, paternalista y bidimensional de la sanidad, constituido por los profesionales sanitarios y el pagador de la sanidad (compañía privada, sistema público o sistema de «tercer sector»), se ha evolucionado –dentro del movimiento *Outcomes*– a un sistema competitivo tridimensional donde se ha incluido la medición científica de la opinión del paciente a la hora de valorar la eficacia de cualquier intervención sanitaria. En esencia podríamos decir:

1) Al pagador de la sanidad le interesará aquello que sea eficaz y de acuerdo a la opinión del comprador de sanidad (el paciente).

2) Al productor de la intervención sanitaria (profesional médico o especialista COT, en nuestro caso) no le quedará más remedio que intentar demostrar que lo que realiza es realmente eficaz.

De esta manera, se ha introducido el uso de instrumentos de medición de la calidad de vida basados en la opinión del paciente dentro de la valoración de resultados en COT. Se ha demostrado que estas medidas científicas de la opinión del paciente son más sensibles a los cambios clínicos tras el tratamiento, que los datos de la exploración física realizados por el propio médico a la hora de medir resultados<sup>5</sup>, opinión que recogiera Hernández Vaquero<sup>6</sup> a la hora de evaluar los resultados de las artroplastias. Estos sistemas de medición de la salud son simples cuestionarios desarrollados, en su mayoría, en EE.UU., Reino Unido o Canadá, por lo que uno de los grandes errores metodológicos en el uso y manejo de estos instrumentos de salud es la traducción por parte del médico o investigador y su aplicación inmediata en una población estudio, sin tener en cuenta los sesgos socioculturales de la traducción, y sin analizar la consistencia interna y reproducibilidad de las mediciones realizadas con dichos instrumentos.

La idea de resultado final no es nueva. Fue introducida por un cirujano ortopédico, Codman (1934), con el propósito de mejorar la eficiencia del cuidado de los pacientes<sup>1</sup>. La idea de resultado final de Codman quedó relegada durante gran parte del siglo XX por varios motivos. Por una parte, la disponibilidad de enormes recursos económicos y el entusiasmo depositado en supuestos avances médicos restaron interés al estudio del coste y la eficacia real de estos últimos. Por otra parte, la aparición de nuevos métodos de diagnóstico y tratamiento dificultó la comparación de los resultados obtenidos, en un mismo proceso, con diferentes sistemas de tratamiento y evaluación<sup>7-10</sup>.

La mejora de los sistemas de información, la voluntad política y social de contener el gasto sanitario y una cierta preocupación por la metodología y la demostración de evidencias científicamente constatadas han reavivado durante las últimas dos décadas el interés de la comunidad médica en general<sup>11</sup> y la ortopédica en particular<sup>12,13</sup> por el estudio

de los resultados finales. En el campo de la Cirugía Ortopédica, este «movimiento de los *outcomes*» se ha visto potenciado por la conjunción de tres circunstancias<sup>14</sup>:

- 1) La escasa calidad metodológica de gran parte de la literatura publicada.
- 2) El estudio de las variaciones de la práctica médica.
- 3) La necesidad de justificar el coste de los procedimientos sanitarios.

Hace más de 70 años Codman estableció el concepto de *outcome* que radicaba en comprender qué resultados producían los tratamientos médicos y quirúrgicos en la recuperación y desarrollo de las actividades diarias de los pacientes. Sin embargo, desde hace tan sólo unos 15 años la comunidad ortopédica, especialmente la anglosajona, viene participando de forma activa en este llamado movimiento-valoración de resultados. Gartland<sup>15</sup>, una de las figuras principales de dicho movimiento, puso de manifiesto las deficiencias evidentes en la bibliografía médica a nuestro alcance, lo que ha dado lugar a una base carente de solidez científica a partir de la cual debíamos tratar a nuestros pacientes. Durante los últimos 10 años el *Outcomes Committee of the American Academy of Orthopaedic Surgeons* (Comité de Resultados de la Academia Americana de Cirujanos Ortopédicos) viene ofreciendo cursos cada año sobre las técnicas empleadas en la evaluación de resultados.

Debemos comprender claramente los diferentes temas que intervienen en la evaluación de los resultados clínicos y funcionales<sup>12,13</sup>. Los resultados clínicos corresponden a lo que tradicionalmente medimos en COT, es decir, los denominados *parámetros clínicos*: tasas de infección, número de luxaciones, amplitud de movimiento, fuerza, marcha, etc. Frecuentemente, estos parámetros se combinan para establecer escalas de valoración (por ejemplo puntuación de la cadera de Harris). Los parámetros clínicos son útiles, pero no siempre presentan una correlación perfecta con el grado de bienestar del paciente. Generalmente, los pacientes que presentan menor rango de movilidad articular se encuentran peor que los que tienen un rango de movilidad completo,

pero no siempre es así. Más aún, aunque nuestro tratamiento consiga restablecer un rango de movilidad completo, eso no asegura el restablecimiento del bienestar del paciente.

Los instrumentos de valoración del estado de salud representan un intento de determinar con mayor precisión la percepción que cada persona tiene de su bienestar o malestar independientemente de los parámetros clínicos que el médico determine. En nuestro campo de acción, la Cirugía Ortopédica, estos instrumentos potencialmente permiten determinar<sup>16,17</sup>:

- 1) En qué grado un proceso concreto (por ejemplo la coxartrosis) se traduce en malestar.
- 2) Cómo una acción médica (por ejemplo, una artroplastia de cadera) modifica el estado de salud del paciente.

Gran parte de la literatura ortopédica publicada antes de la década de los noventa presenta escasa calidad metodológica, situación sobre la que autores como Gartland<sup>15</sup> y Gross<sup>18</sup> llamaron la atención acerca de la inconsistencia metodológica de la literatura hasta entonces publicada sobre la artroplastia total de cadera. Actualmente resulta necesario demostrar que los procedimientos terapéuticos que ofrecemos a los pacientes conducen a mejoras significativas de su calidad de vida.

El primer requisito para valorar el impacto de un determinado procedimiento sobre la calidad de vida es utilizar instrumentos validados para la población a estudiar, teniendo en cuenta variaciones culturales y lingüísticas. Estos datos se obtienen del paciente mediante el uso de cuestionarios validados y se centran en el desarrollo de las funciones y actividades de cada paciente como un todo, partiendo desde el punto de vista de éste. Este tipo de evaluación se ocupa de todos los campos de la función humana, tales como la función física, el dolor, el estado de salud, la actividad laboral y las actividades de la vida diaria (fig. 1).

Los cuestionarios se desarrollan de acuerdo a unos principios psicométricos establecidos. Los pasos que se siguen son el desarrollo de las cuestiones que se van a preguntar, la selección del formato, la simplificación de las preguntas y

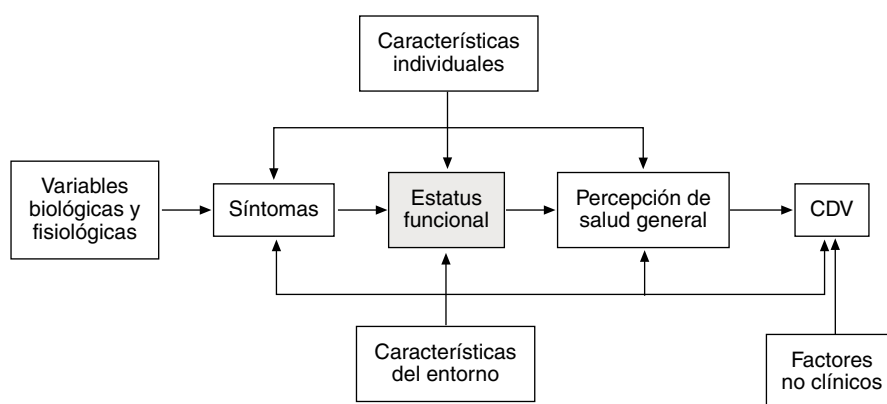


Figura 1. Medición de resultados en modelo de calidad de vida (CDV) relacionada con la salud.

la validación. Las cuestiones o preguntas se obtienen a través de expertos en el tratamiento del proceso que se evalúa, mediante otros cuestionarios o, como sería lo ideal, a través de los pacientes que sufren el problema o problemas que se están estudiando. El cuestionario puede tener un formato de respuesta SÍ/NO (o aprobatorio), un formato de escala (escala Likert de 3 a 5 puntos con opiniones finales) o un formato de escala analógico visual donde el paciente hace una marca en un segmento lineal establecido. Cada uno de estos formatos tiene ventajas e inconvenientes. La simplificación de las preguntas supone de forma sistemática una disminución en el número de puntos a tratar. Se basa en la frecuencia con la que el paciente que es objeto de estudio, pone de relieve el tema y el impacto relativo de la cuestión (tabla 1). Las técnicas estadísticas suelen proporcionar una amplia gama de cuestiones con diversos efectos sobre la calidad de vida y el desarrollo de sus actividades diarias.

La validación abarca la validez patente, la validez de constructo y la validez de criterio, además de suponer una prueba de fiabilidad y de receptividad. La validez patente de un cuestionario responde sencillamente a la pregunta de si éste resulta razonable para la condición o condiciones para las que se diseñó el estudio. La validez de constructo implica la recogida de datos, el desarrollo de hipótesis acerca de cómo deberían ser las respuestas a este cuestionario, y la evaluación de si estas respuestas resultan consistentes para dichas hipótesis. La validez de criterio implica la recogida simultánea de datos junto con el uso de un modelo establecido para determinar cómo se relacionan los datos recogidos a través del nuevo cuestionario con los datos estándar. La fiabilidad se comprueba mediante la repetición del cuestionario, ya que las respuestas a las preguntas deberían permanecer estables si el estado clínico del paciente no ha cambiado. Por último, la receptividad es un ejercicio que se lleva a cabo para evaluar la capacidad que posee el cuestionario de determinar los cambios en el estado funcional del paciente.

En pocas palabras, el cuestionario debe parecer razonable tanto a los pacientes como a los profesionales médicos, y se debe comprobar que mide aquello para lo que estaba diseñado de un modo predecible. A la hora de elegir cuál es de los cuestionarios validados el más conveniente, hemos de centrarnos en el propósito de la recogida de datos, los recursos de que disponemos y otras cuestiones de factibilidad.

Existen dos grandes grupos de instrumentos de valoración del estado de salud: genéricos y específicos<sup>19</sup>. Los instrumentos *genéricos* están diseñados para evaluar el estado de salud en cualquier población de pacientes independientemente de sus características poblacionales o del tipo de enfermedad que presenten. Existen dos tipos de instrumentos genéricos: los perfiles de salud y las medidas de utilidad. Los perfiles de salud intentan medir todos los aspectos importantes de la calidad de vida relacionada con la Salud (CVRS) en varias dimensiones (física, psicosocial) y cate-

**Tabla 1.** Preguntas clave en la medición de resultados

Contenido, población/escenario, objetivo
¿Qué resultados queremos medir?
¿Qué población queremos estudiar y en qué escenario?
¿Cuál es el objetivo del estudio? ¿describir, predecir o medir un cambio?
Validez del contenido
¿Qué dominios y elementos están incluidos?
¿Existen omisiones importantes o inclusiones inapropiadas?
Diseño válido
¿Cada pregunta está formulada de forma fácil de entender?
¿Son las categorías de respuesta correctas?
¿Existe una cuantificación que sume las preguntas? ¿Cómo se calcula?
Aplicabilidad
¿Es fácil de entender?
¿Es fácil de usar? (por ejemplo manual de instrucciones claro)
¿Es aceptable?
¿En qué formato está disponible? (autorrespuesta, teléfono, entrevista)
¿Cuánto tarda en contestarse?
¿Quién realiza el cuestionario?

gorías (trabajo, sueño, etc.). Las medidas de utilidad son puntuaciones que reflejan tanto el estado de salud del paciente como el valor de dicho estado de salud para el paciente, representando el impacto neto sobre la cantidad y la calidad de vida del paciente. Los instrumentos *específicos* están diseñados para valorar el estado de salud en un tipo concreto de enfermedad (artritis reumatoide), población (ancianos frágiles), función (deambulación) o problema (dolor). Los instrumentos genéricos y específicos presentan diferentes ventajas e inconvenientes (tabla 2). El principal atractivo de los genéricos es la posibilidad de comparar el impacto relativo de diferentes programas de salud (por ejemplo, se puede comparar la mejora de la calidad de vida proporcionada por la artroplastia de cadera y el trasplante cardíaco). Sin embargo, pueden resultar menos sensibles al cambio. Los instrumentos específicos tienen como principal ventaja ser más discriminativos, presentar mayor fiabilidad y respuesta al cambio; su inconveniente es que no permiten la comparación entre diferentes poblaciones o procesos. Ejemplos de cuestionarios validados de mayor aplicación en nuestra especialidad se muestran en la tabla 3.

Lo ideal es emplear simultáneamente un cuestionario genérico –que mida la salud global– y otro específicamente designado para un tipo de paciente o patología concretos. El cuestionario genérico de elección en la actualidad es el SF-36, validado en castellano<sup>20,21</sup>. Los cuestionarios WOMAC, MFA y MODEMS probablemente sean los específicos de mayor aplicación en nuestro campo en el futuro. El proyecto MODEMS tiene fundamentalmente dos metas. La primera es la de crear una base de datos nacional (EE.UU.) sobre los resultados de los tratamientos musculoesqueléticos, y la segunda es la de proporcionar un método mediante el cual los médicos puedan valorar los resultados de los tratamien-

**Tabla 2.** Características de los instrumentos de valoración de la calidad de vida relacionada con la salud según Guyat<sup>17</sup>

	Ventajas	Inconvenientes
Genéricos	Instrumento único	Puede que no se centre adecuadamente en el área de interés
Perfil de salud	Detecta efectos diferenciales sobre diferentes aspectos del estado de salud Es posible la comparación entre intervenciones y enfermedades	Puede que no responda al cambio
Medida de utilidad	Puntuación única que refleja el impacto neto sobre la cantidad y calidad de vida Permite análisis coste-utilidad Incorpora la muerte	Su determinación es difícil No permite analizar el efecto sobre diferentes aspectos de la calidad de vida Puede que no responda al cambio
Específicos	Clínicamente interpretable Puede tener mayor respuesta al cambio	No permite comparaciones entre intervenciones o enfermedades Su aplicación puede verse limitada a poblaciones de pacientes muy concretas

tos musculoesqueléticos en pacientes externos y compararlos con los de la media nacional<sup>22-25</sup>.

En España, la Agencia de Evaluación de Tecnologías Sanitarias<sup>26</sup> editó los «Índices y Escalas utilizados en ciertas tecnologías de la prestación ortoprotésica (Protetización del Sistema Osteoarticular)», en donde se analizan los cuestionarios genéricos y específicos de aplicación en nuestra especialidad, así como sus aplicaciones, ventajas e inconvenientes. La aplicación de estos instrumentos ya ha comenzado a proporcionar datos interesantes en el campo de la Traumatología y Cirugía Ortopédica. De acuerdo con estos datos, las artroplastias de cadera o rodilla y varios procedimientos quirúrgicos sobre la columna mejoran significativamente la calidad de vida de los pacientes, mientras que los traumatismos de alta energía afectan negativamente la salud durante dos o más años, asimismo la mejor calidad

de vida de los pacientes que conservan un miembro inferior afectado por una neoplasia o un traumatismo importante.

Tras la evaluación de los cuestionarios disponibles y de contemplar diferentes enfoques en la valoración de los resultados obtenidos por los tratamientos en pacientes que padecían trastornos musculoesqueléticos, la AAOS y las sociedades especializadas establecieron un marco para la creación de 4 cuestionarios que cubrirían el espectro de trastornos musculoesqueléticos:

- 1) Enfermedades y lesiones pediátricas.
- 2) Enfermedades y lesiones de la columna vertebral en adultos.
- 3) Enfermedades y lesiones de las extremidades superiores en adultos.
- 4) Enfermedades y lesiones de las extremidades inferiores en adultos.

La AAOS y las Sociedades coincidieron tanto en que los cuestionarios debían tomar como base principal los síntomas, la recuperación de las funciones básicas y la calidad de vida, como en que debían ser fidedignos, validados y susceptibles de cambios en la función musculoesquelética. En muchos casos los 4 instrumentos incluían componentes de cuestionarios anteriores. Los primeros cuestionarios fueron distribuidos para su evaluación en febrero de 1995 y últimamente se ha incorporado un instrumento de evaluación musculoesquelético breve y genérico, el *Short Musculoskeletal Functional Assessment* (SMFA), que abarque las características de la columna vertebral y las extremidades inferiores de adultos en un único cuestionario.

La valoración crítica de los resultados del tratamiento es uno de los métodos más importantes y eficaces en la mejora de la atención sanitaria. La información válida y fidedigna acerca de los resultados o consecuencias del tratamiento permite identificar los tratamientos más eficaces y tomar decisiones con fundamento en lo referente a la asignación de los recursos sanitarios. Aunque es evidente que la

**Tabla 3.** Cuestionarios validados de mayor aplicación en Cirugía Ortopédica<sup>2</sup>

Genéricos	
QWB	<i>Quality of Well-Being Scale</i>
SIP	<i>Sickness Impact Profile</i>
HIE	<i>Health Insurance Experiment</i>
MHIQ	<i>McMaster Health Index Questionnaire</i>
SF-36	<i>Short-Form 36</i>
NHP	<i>Nottingham Health Profile</i>
DHP	<i>Duke Health Profile</i>
QL-Index	<i>Quality of Life Index</i>
HUI	<i>Health Utility Index</i>
EuroQol	<i>EuroQol</i>
WHOQOL	<i>WHOQOL</i>
Específicos	
WOMAC	<i>Western Ontario and McMaster University Osteoarthritis Index</i>
MFA	<i>Musculoskeletal Functional Assessment</i>
MODEMS	<i>Musculoskeletal Outcomes Data Evaluation and Management System</i>

información obtenida con estos instrumentos resulta de gran interés, aún quedan incógnitas importantes por despejar. Hay quien vaticina que el movimiento de los resultados finales será pasajero y no supondrá una modificación sustancial del quehacer médico. Aunque lo ideal sería analizar si la actividad asistencial de cada uno de nosotros realmente mejora la calidad de vida de nuestros pacientes, existen obstáculos importantes para la introducción de estos instrumentos en la práctica cotidiana. Su utilización sistemática conlleva mayores recursos económicos y laborales. Por el momento, el colectivo de especialistas en Ortopedia parece decidido a seguir ampliando sus redes de recogida de datos sustanciales en los ámbitos de la práctica médica, de cada estado y de las subespecialidades. Nuestros pacientes recibirán un servicio mejor si trabajamos juntos en la identificación de las cuestiones clínicas más relevantes y de las fuentes de experiencia para coordinar mejor la investigación.

En el caso de la Cirugía Ortopédica española, la necesidad de repetir el proceso de validación en nuestro idioma supone un obstáculo adicional. Por todo ello, es posible que en nuestro país la utilización de estos instrumentos quede momentáneamente restringida al campo de la investigación clínica. En la docencia, la utilidad práctica de las pruebas, su sensibilidad y especificidad, son un camino dentro del binomio enseñanza-aprendizaje.

En esta Comunicación Oficial del Congreso SECOT en Sevilla, queremos:

1) Transmitir los conocimientos del «movimiento-valoración de resultados» en nuestra especialidad.

2) Manifiestar nuestra inquietud por la implantación y desarrollo en España de los instrumentos de valoración del estado de salud y su aplicación a la evaluación de resultados en COT.

3) Facilitar a nuestros especialistas:

a) La COT basada en la evidencia ante la valoración de resultados. Su búsqueda.

b) La VR en el miembro superior, en el miembro inferior y en la columna vertebral.

## **LA CIRUGÍA ORTOPÉDICA Y TRAUMATOLOGÍA BASADA EN LA EVIDENCIA ANTE LA VALORACIÓN DE RESULTADOS**

La mejora en el acceso al documento es una de las principales aportaciones de internet en el ámbito bibliotecario y de la información en general, y una de las más perceptibles por parte del usuario final, que ya no necesita asistir físicamente a la biblioteca o centro de documentación. Ello viene a resolver una de las barreras de accesibilidad a la información definidas por Jovell<sup>27</sup>, la falta de tiempo, pero no solventa la falta de interés, recursos y conocimientos.

La falta de interés se asocia con frecuencia a una menor capacidad para utilizar la evidencia actualizada sobre aten-

ción sanitaria en la práctica de la medicina<sup>28</sup>, es lo que se denomina una «ignorancia evitable»; aspectos como la barrera idiomática es una de las más graves. Sotomayor<sup>29</sup>, editor de la Revista Médica de Santiago (electrónica) reivindica la necesidad de editar revistas en castellano para los 900.000 médicos hispanoparlantes de América Latina donde tan sólo una minoría domina suficientemente el inglés como para leer fluidamente la literatura; señala que debemos publicar en la lengua materna, entonces ésta no desaparecerá y habrá verdadero desarrollo científico. Pero la realidad es que el idioma científico internacional es el inglés y hemos de dominar el inglés aunque sea el «pasivo», con disciplina y diccionario.

Las otras dos barreras, la carencia de recursos y conocimientos debieran de ser fácilmente franqueables precisamente por la cómoda accesibilidad que ha supuesto internet.

Mas la escasez de recursos como los ordenadores y la falta de adaptación a las nuevas tecnologías y metodologías informáticas sigue siendo un freno a la información en línea que no iguala a la información clásica en uso, en gran parte coloquial<sup>30</sup>.

Una vez enfrentados al ordenador la obtención de la información es «fácil» y esto ¿es una ventaja? Más que una ventaja realmente ha creado una barrera mayor por diferentes aspectos que podemos agrupar, según el problema radique, en la cantidad o en la calidad de los trabajos.

En relación con la cantidad cabe destacar el problema del almacenamiento y exceso de la información. Tal es la cantidad y dispersión de la información que se mueve en la red que hace muy difícil decidir cuál es la más conveniente para resolver la duda que tenemos; accederemos a la misma mediante buscadores generales (Google, Yahoo, Lycos, etc.), buscadores y portales médicos (E-medicine, Medscape, etc.), portales específicos de Cirugía Ortopédica y Traumatología (Orthogate, Orthoguide, etc.), pero los accesos más fiables que agrupan a numerosas revistas científicas son las bases de datos, la más relevante y usada es *MEDLINE* a la que llegaremos a través de *PubMed*.

*Pubmed* (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed>) es un proyecto desarrollado por la *National Center for Biotechnology Information (NCBI)* en la *National Library of Medicine (NLM)* que permite el acceso a bases de datos bibliográficas compiladas por la *NLM*. *MEDLINE*, antiguo *Index Medicus*, es la base de datos más importante de la *NLM* abarcando los campos de la Medicina, Oncología, Enfermería, Odontología, Veterinaria, Salud Pública y Ciencias preclínicas. Actualmente contiene más de 15 millones de referencias bibliográficas de artículos de revistas desde el año 1950, provenientes de 4.780 revistas internacionales de Ciencias de la Salud (datos del 2004).

La inmediatez de la información se fomentó con la creación en 1996 de *PREMEDLINE*, una base de datos que suministra datos básicos y resúmenes antes de que se añadan todos los campos en los registros completos y sean incorpo-

rados a Medline. Cada día se introducen nuevos registros a Premedline, más de 30.000 al mes (Hubbs 1998) y cada registro recibe un número de identificación PMID. Una vez que se les incorpore los términos MeSH y otros datos de la indexación, se suprimen de Premedline y se incorporan en Medline. Tanto en una base como en la otra los artículos aparecen incluso antes de ver la luz en formato papel.

Es fácilmente comprensible que tal cantidad de información en lugar de ayudar a la toma de decisiones la dificulte aún más; valga la reflexión de que si leyéramos dos artículos diarios de los 6 millones de artículos médicos que se publican en un año, necesitaríamos 82 siglos para poderlos leer. Sin embargo, la demanda de los mismos se concentra enormemente, pues un 20% de los artículos de revista satisface el 80% de la demanda y un 50% de los artículos publicados, no se citan jamás<sup>31</sup>.

A pesar de todo se considera Medline como la mejor fuente que conocemos para buscar los datos necesarios para resolver problemas en medicina (McKibbin 1994), aunque autores como Pritchard<sup>32</sup>, era escéptico sobre que Medline pudiera ser una buena herramienta de asesoramiento médico (para él la utilidad de las búsquedas automatizadas sería escasa, precisaría una gran cantidad de tiempo y además la pertinencia y la validez de lo recuperado es escasa. Así decía que «La base de datos Medline es de ayuda para búsquedas de literatura médica sin prisas, pero para obtener respuestas a problemas clínicos importantes es virtualmente inútil»). A esta aseveración llegaba tras realizar búsquedas automatizadas para responder a preguntas generadas en la práctica; menos del 1% de los artículos encontrados contribuían a cambiar el tratamiento de los pacientes.

Una vez identificados los artículos deseados surge un nuevo obstáculo, acceder a los mismos a texto completo, dado que la mayoría de éstos no son gratuitos (hecho paradójico al estar en una red diseñada para la difusión gratuita de la información y movimiento de acceso libre <http://www.earlham.edu/~peters/fos/guide.htm#s->)

Solventado el problema de obtener los trabajos aparece el escollo de determinar si son relevantes o no, pues existe mucha información anecdótica, en ocasiones relevante; incorporan con frecuencia las opiniones del autor y adolecen de los sesgos de publicación. Pero ¿qué trabajos son de mejor calidad?

En 1976 el Ministerio de Sanidad Canadiense constituyó una comisión que habría de analizar un conjunto de medidas sanitarias y recomendarlas como beneficiosas para toda la población de dicho país. Para ello desarrolló una metodología encaminada a adjudicar unos pesos a la evidencia científica suministrada por el análisis de la bibliografía sobre la efectividad de cada intervención sanitaria. Fruto de ello fue la elaboración por la *Canadian Task Force on the Periodical Health Examination*, adaptado con posterioridad por la *US Preventive Task Force* de un sistema de jerarquización o graduación de la calidad de la evidencia. Ge-

neraron los niveles de evidencia para jerarquizar la validez de ésta sobre el valor de actuaciones preventivas, y los ligaron a los llamados «grados de recomendación». La mayor parte de las escalas descritas coinciden en calificar con más credibilidad a aquellos estudios aleatorios, controlados y prospectivos. Nosotros proponemos la escala utilizada en la Agencia de evaluación de Tecnología Médica, que es adaptación de la desarrollada por Jovell<sup>27</sup> (tabla 4).

Esta iniciativa canadiense fue liderada por Sackett en 1992<sup>33</sup>, año en que publica por primera vez el término de medicina basada en la evidencia. Así se inicia un cambio del paradigma antiguo en el que los médicos resuelven los problemas clínicos a los que se enfrentan reflexionando sobre su propia experiencia o sobre la fisiopatología subyacente, consultando un libro de texto, leyendo las secciones de introducción y discusión de un artículo relevante procedente de una revista actual o preguntando a un experto local. Este paradigma confiere un elevado valor a la autoridad científica tradicional, y a menudo las respuestas tratan de encontrarse en lo publicado por expertos locales o internacionales. Se inicia el camino a un nuevo paradigma basado en la información deducida de la experiencia clínica, más de la colectiva que de la personal, y de la obtenida tras consultar regularmente la literatura original y tras analizarla de forma crítica. Este nuevo paradigma confiere un valor menor a la autoridad<sup>34</sup>.

En esencia la medicina basada en la evidencia (MBE) aporta un marco conceptual nuevo para la resolución de problemas clínicos concretos mediante un proceso de búsqueda sistemática, evaluación crítica y aplicación de los hallazgos de la investigación a la toma de decisiones clínicas; es un medio para afrontar la existencia de una enorme información científica y en evolución continua, la exigencia de ofrecer la máxima calidad asistencial y la limitación de recursos destinados a la atención sanitaria; pero no resuelve la ausencia de investigación básica sobre un tema. Es un método de trabajo que se puede considerar de excelencia<sup>35</sup>, al combinar en una el concepto de alta calidad documental, la unificación de las fuentes de información y el acceso a éstas mediante artículos de texto completo. El mejor ejemplo de esta metodología es la Biblioteca Cochrane plus (<http://www.update-software.com/clibplus/clibplus.asp>) a la que podemos acceder de forma gratuita por la página del Ministerio de Sanidad y Consumo ([http://www.msc.es/Diseno/informacionProfesional/profesional\\_biblioteca.htm](http://www.msc.es/Diseno/informacionProfesional/profesional_biblioteca.htm)).

Pero este halo de bondad que poseen bibliotecas y revistas que se nutren esencialmente de metaanálisis no debe deslumbrarnos, ya que la búsqueda inmediata y personal de la información se ha de seguir realizando regularmente, aunque con los mismos criterios que rigen a la MBE, pues, si no, seremos clínicos «pseudoinformados» fruto de una deficiente recogida de información, una lectura incompleta de la misma y una interpretación incorrecta. Una de las fuentes más importantes de «pseudoinformación pseudo-

**Tabla 4.** Escala utilizada en la Agencia de Evaluación de Tecnología Médica que es adaptación de la desarrollada por Jovell<sup>27</sup>

Escala de Evaluación de la Evidencia de la Agencia de Tecnología Médica				
Niveles	Calidad de la evidencia	Tipo de diseño del estudio	Condiciones de rigor científico	Recomendación
I	Buena	Metaanálisis de ensayos controlados y aleatorios	Análisis de datos de pacientes individuales Metarregresión Diferentes técnicas de análisis Ausencia de heterogeneidad Calidad de los estudios	Existe adecuada evidencia científica para recomendar o desaconsejar la adopción de la tecnología
II	Buena	Ensayos controlados y aleatorios de muestra grande	Evaluación del poder estadístico Calidad del estudio multicéntrico	Existe adecuada o cierta evidencia científica para recomendar o desaconsejar la adopción de la tecnología
III	Buena	Ensayos controlados y aleatorios de muestra pequeña	Evaluación del poder estadístico Calidad del estudio multicéntrico	
IV	Regular	Ensayos prospectivos controlados no aleatorios (controles coincidentes en el tiempo)	Evaluación del poder estadístico Calidad del estudio multicéntrico	Existe cierta evidencia científica para recomendar o desaconsejar la adopción de la tecnología
V	Regular	Ensayos prospectivos controlados no aleatorios (controles históricos no coincidentes en el tiempo)	Controles históricos	
VI	Regular	Estudios de control (cohorte)	Calidad del estudio multicéntrico	Existe insuficiente evidencia científica para recomendar o desaconsejar la adopción de la tecnología
VII	Regular	Estudios caso-control	Apareamiento	
VIII	Baja	Series clínicas no controladas Estudios descriptivos: seguimiento de la enfermedad, vigilancia epidemiológica, registros, bases de datos Comités de expertos, conferencias de consenso		
IX	Baja	Anécdotas o casos		

científica» procede de las compañías farmacéuticas. El hecho de que en nuestro medio sean las principales fuentes de soporte financiero de la investigación y la formación de los médicos deriva en una información sesgada, dirigida a intereses comerciales, y en la que lo «científico» y lo «promocional» es a menudo indistinguible<sup>36</sup>.

### ¿CÓMO REALIZAR LAS BÚSQUEDAS BIBLIOGRÁFICAS?

Realizar una búsqueda en *Medline* (<http://www4.ncbi.nlm.nih.gov/entrez/query.fcgi>) resulta fácil, o al menos así lo parece; en la misma página *web* hay un tutorial para aprender a moverse por la base de datos y obtener la información deseada (<http://www.ncbi.nlm.nih.gov/entrez/query/static/help/pmhhelp.html#SearchFieldQualification>)<sup>37</sup> y por tanto no va a ser objeto de este trabajo. Como ejemplo vamos a realizar la búsqueda de artículos con un nivel I de evidencia, ensayos clínicos controlados y aleatorios, de *outcomes* de fractura de cadera.

De hecho, basta con poner una palabra clave (por ejemplo *fracture*) en el campo de búsqueda una vez señalada la

base de datos *Pubmed* y pulsar adelante (*go*) y ya está: ¡132.832 artículos a fecha de 24-05-2005! Si observamos en el campo «Detalles» de la búsqueda (*Details*) se ha realizado una búsqueda matizada con una serie de operadores booleanos y campos definidos («*fractures*»[TIAB] NOT *Medline*[SB]) OR «*fractures*»[MeSH Terms] OR *fracture*[Text Word]). Si el término que buscamos es uno pero con dos palabras (por ejemplo: *hip fracture*) la búsqueda es similar («*hip fractures*»[TIAB] NOT *Medline*[SB]) OR «*hip fractures*»[MeSH Terms] OR *hip fracture*[Text Word]) pero lógicamente con menor número de trabajos: 11.570. Mas si ponemos las palabras entre comillas definimos que la búsqueda ha de ceñirse a esa palabras exactamente y en el orden establecido («*hip fracture*»), la búsqueda se realiza tal como la ordenamos «*hip fracture*»[All Fields]- y el número de publicaciones se reduce a 3.547, que aún son muchas.

Si afinamos la búsqueda y seleccionamos tan sólo los ensayos clínicos y los ensayos controlados aleatorios (por ejemplo «*hip fracture*» «*clinical trial*» «*randomized controlled trial*») el número de artículos encontrados en *Medline* es ya de 176 trabajos y la búsqueda automática ha puesto los operadores booleanos. Por último si queremos saber cuantos de éstos hacen referencia a los *outcome* y añadimos



este término a la búsqueda (Ej: «hip fracture» «clinical trial» «randomized controlled trial» outcome) el resultado se limita a 72 publicaciones y la secuencia definida ha sido «hip fracture»[All Fields] AND «clinical trial»[All Fields] AND «randomized controlled trial»[All Fields] AND outcome[All Fields] AND Randomized Controlled Trial[ptyp].

Llegados a este punto, si estamos conformes con lo hallado, es el momento de acceder a los artículos a texto completo y no conformarnos con los resúmenes. *Pubmed* nos muestra unas carpetas con unas franjas verde, o naranja y verde que nos indican cuáles de ellos están disponibles gratuitamente.

Realmente lo que hemos realizado ha sido una estrategia de búsqueda y no basta con hacerla, sino con hacerla bien, de tal modo que el número de artículos encontrados sea de forma adecuada a la solicitud de calidad realizada. Esto supone que a medida que la búsqueda es más compleja y especializada es necesario conocer técnicas de recuperación de información, sobre todo cuando la consulta se está realizando en bases de datos bibliográficas. Un mayor conocimiento de las fuentes, sistemas y técnicas de recuperación implicará mejores resultados, obtenidos más rápidamente, lo que supone más calidad y mayor ahorro, tanto de tiempo como de dinero, pero eso no es factible para un gran número de usuarios de las bases de datos.

Heersh y Hickman<sup>30</sup>, tras realizar un análisis de diferentes métodos de búsqueda y el uso que de ellos se realizaba, concluyeron en que las estrategias de búsqueda son sensiblemente más efectivas que los métodos simples de búsqueda, consiguiendo disminuir el número de artículos a leer<sup>38</sup>.

En los últimos años se han propuesto distintas estrategias, algunas de ellas orientadas a especialidades médico-quirúrgicas concretas, tales como Anestesia<sup>39</sup>, Medicina de Familia<sup>40</sup>, Ortopedia<sup>41,42</sup> y por supuesto a la Cirugía Ortopédica<sup>43</sup>.

Sin embargo, para que estas estrategias sean válidas es preciso definir previamente un patrón oro. Como tal se define el número de artículos coincidentes con las variables establecidas encontrados mediante búsqueda manual.

Contrastados los resultados de la estrategia motorizada en Medline con la manual se establecen los criterios de<sup>38</sup>:

1) Sensibilidad: relaciona el número de artículos del patrón oro encontrados con el número total de artículos del patrón.

2) Precisión: relaciona el número total de artículos patrón con el número total de artículos encontrados.

3) NNR (número de referencias irrelevantes) que hay que buscar para encontrar una relevante, es el cociente inverso de la precisión.

Haynes<sup>44</sup> sigue la siguiente formulación (tabla 5) para poder calcular estos parámetros.

Así se han validado estrategias para seleccionar diferentes tipos de estudio orientados al tratamiento, pronóstico, diagnóstico y etiología y diferentes diseños<sup>45,46</sup> o sólo al

Tabla 5

Términos de búsqueda	Revisión manual	
	Cumplen el criterio	No cumplen el criterio
Detectados	a	b
No detectados	c	d
	a+b	c+d

Sensibilidad:  $a/a+c$ ; especificidad:  $d/(b+d)$ ; precisión:  $a/a+b$ ; exactitud:  $(a+d)/(a+b+c+d)$ ; todos los artículos clasificados en la revisión manual =  $a+b+c+d$ .

diagnóstico<sup>40</sup> empleando diferentes términos siendo común a todas ellas los términos de sensibilidad, especificidad y precisión (*accuracy*).

Una forma de mejorar la calidad y precisión de la búsqueda es seleccionar un número determinado de revistas clínicas acreditadas por su calidad en lo publicado, pues la precisión de las búsquedas depende de la concentración de artículos relevantes en la base de datos, pero Medline contiene muchas revistas no clínicas y por consecuencia la concentración de artículos de tratamiento con alta calidad será menor que en la totalidad de Medline, y la precisión de la búsqueda será menor que la esperada<sup>44</sup>.

La estrategia de búsqueda más utilizada es la de Haynes<sup>44,45</sup> que se incorpora al motor de búsqueda automático de Pubmed en el apartado de *Clinical Queries*, es por tanto un buscador especializado que tiene incorporados «filtros metodológicos» para búsquedas clínicas y revisiones sistemáticas.

Tiene una opción de *Systematic Reviews* que localiza los artículos de revisiones sistemáticas y metaanálisis, revisiones de ensayos clínicos, conferencias de consenso, guías de práctica clínica, etc.

Pero la gran utilidad que aporta es la búsqueda por filtros. Propone 4 categorías: terapéutica (*therapy*), diagnóstico (*diagnosis*), etiología (*etiology*) y pronóstico (*prognosis*). Pone énfasis sobre dos categorías: sensibilidad (*sensitivity*), este toma artículos relevantes y no tan relevantes y especificidad (*specificity*), que es más precisa pero con menor cantidad de artículos. Basta con introducir las palabras clave sin necesidad de poner límites, operadores booleanos, campos, etc. y marcar las opciones de búsqueda y el motor establece la estrategia de búsqueda a seguir. Según sean las opciones elegidas accederemos a un porcentaje de artículos que cumplan el nivel de evidencia seleccionado (tabla 6).

Volviendo al ejemplo anterior introducimos en el campo de búsqueda los términos *hip fracture* y *outcome*, marcamos las opciones de *therapy* y *narrow specific search* y tras lanzar la búsqueda obtenemos 75 estudios.

Pero estas estrategias presentan aún deficiencias como:

1) Muchos trabajos presentan errores metodológicos importantes, lo que dificulta su clasificación.

Tabla 6. Búsqueda por filtros

Category	Optimized For	Broad/Narrow	PubMed Equivalent
Therapy	Sensitive/broad	99%/70%	(clinical[Title/Abstract] AND trial[Title/Abstract]) OR clinical trials[MeSH Terms] OR clinical trial[Publication Type] OR random*[Title/Abstract] OR random allocation [MeSH Terms] OR therapeutic use[MeSH Subheading])
	Specific/broad	93%/97%	(randomized controlled trial[Publication Type] OR (randomized[Title/Abstract] AND controlled[Title/Abstract] AND trial[Title/Abstract]))
Diagnosis	Sensitive/broad	98%/74%	(sensitiv*[Title/Abstract] OR sensitivity and specificity[MeSH Terms] OR diagnos*[Title/Abstract] OR diagnosis[MeSH noexp] OR diagnostic + [MeSHnoexp] OR diagnosis, differential[MeSHnoexp] OR diagnosis[Subheading:noexp])
	Specific/narrow	64%/98%	(specificity[Title/Abstract])
Etiology	Sensitive/broad	93%/63%	(risk*[Title/Abstract] OR risk*[MeSH no exp] OR risk * [MeSH no exp] OR cohort studies [MeSH Terms] OR group*[Text Word])
	Specific/narrow	51%/95%	(relative[Title/Abstract] AND risk*[Title/Abstract]) OR (relative risk[Text Word]) OR risks [Text Word] OR cohort studies[MeSHno exp] OR (cohort[Title/Abstract] AND stud*[Title/Abstract]))
Prognosis	Sensitive/broad	90%/80%	(incidence[MeSH:no exp] OR mortality[MeSH Terms] OR follow up studies [MeSH:noexp] OR prognos*[Text Word] OR predict*[Text Word] OR course*[Text Word])
	Specific/narrow	52%-94%	(prognos*[Title/Abstract] OR (first[Title/Abstract] AND episode[Title/Abstract]) OR cohort [Title/Abstract])
Clinical prediction guides	Sensitive/broad	96%/79%	(predict*[tiab] OR predictive value of tests[mh] OR scor*[tiab] OR obsrev*[tiab] OR observer variation [mh])
	Specific/narrow	54%/99%	(validation[tiab] OR validate[tiab])

2) Se indexan términos incoherentes.

3) Se indexan términos y métodos que son modificados en el tiempo y en ocasiones se realizan cambios retrospectivamente.

4) Muchos términos tienen el mismo significado.

5) No son extrapolables a *Embase* por existir diferencias en el indexado de la base de datos.

6) Los filtros para estudios diagnósticos no son plenamente satisfactorios porque las búsquedas sensitivas tienen poca precisión.

De ahí que tengan que estar revisándose y mejorándose periódicamente tal y como realiza el grupo de Haynes. En la última revisión realizada por este grupo, en la que introducen términos de economía, peritaje, supervivencia, etc., consiguen una mejora en la selección de estudios clínicos controlados, aleatorios o de doble ciego, comparando los resultados de búsqueda obtenidos con 19 estrategias de búsqueda publicadas con anterioridad.

Para finalizar queremos hacer una reflexión sobre la calidad de la información de nuestra especialidad que se publica.

## ¿QUÉ CALIDAD TIENEN LOS TRABAJOS DE CIRUGÍA ORTOPÉDICA Y TRAUMATOLOGÍA PUBLICADOS?

Es difícil contestar esta pregunta por diferentes motivos como:

1) Los ensayos clínicos en nuestra especialidad están muy limitados por la ética del modelo de investigación.

2) La mayor parte de los estudios que se realizan son de cohorte o de revisión, y la evidencia en éstos baja sensiblemente.

3) Los artículos de consenso a los que muchos profesionales dan una gran importancia son considerados de baja evidencia.

Existe una tendencia en los investigadores a publicar en las revistas de mayor índice de impacto y las de un impacto muy alto no son específicas de Cirugía Ortopédica y Traumatología (COT). Ello implica que la búsqueda y selección de trabajos hay que realizarla en el Medline completo, lo cual disminuye la exactitud de la misma.

Algunos autores han querido analizar la calidad de los artículos publicados en revistas de COT.

Rudicel y Esdale<sup>47</sup> realizaron una búsqueda en Medline de ensayos controlados aleatorios en cirugía ortopédica en que por lo menos uno de los procedimientos comparados fuera quirúrgico, y ellos encontraron sólo 4 estudios publicados durante los años 1979 a 1984.

Kiter<sup>48</sup> revisó el número de ensayos controlados aleatorios y metaanálisis publicados en 9 revistas de COT<sup>27</sup> recogidas en Medline entre los años 1966 y 1999, y observan que en 33 años se habían publicado 671 ensayos controlados aleatorios y 12 metaanálisis, estando el 81,9% de los ensayos publicados en 4 revistas<sup>27</sup> y el 66% de ellos se refirieron a artroplastias.

Zanoli<sup>49</sup> amplía el estudio a 44 revistas detectando cómo el número de publicaciones con una evidencia alta por su metodología ha aumentado ostensiblemente a partir de 1993, si bien los trabajos que hacen referencia a la MBE son aún muy pocos.

Esta mejora en la calidad de los artículos en la década de los noventa se observa también en el resto de las ramas médicas, existe en la red un incremento del interés en estu-

dios de resultados de medidas basadas en el paciente, particularmente ensayos controlados aleatorios, es lo que se denomina POEMS (*Patient-oriented Evidence That Matters*)<sup>50</sup>. Es tal interés que ha aparecido una página web (INFOPOEMS) dedicada a recoger éstos (<http://www.info-poems.com/>).

De lo que se deduce que, en general, el grado de evidencia real que soporta nuestra práctica clínica es bastante endeble. Bhandari<sup>51</sup> revisó los trabajos publicados en el año 2000 de 15 revistas de cirugía ortopédica con factor de impacto y analizó el índice de citación que tuvieron; tras el análisis de los mismos concluyó que las revisiones sistemáticas rigurosas alcanzaban un mayor número de citaciones que el resto de los trabajos; el rigor de las revisiones fue un predictor significativo del número de citaciones en otras revistas, o lo que es lo mismo, a mayor calidad mayor difusión.

Por lo que es de agradecer todo tipo de mejoras encaminadas a mejorar la calidad de las publicaciones, tales como:

1) La introducción de los criterios de MBE en las revistas; así el *Journal Bone and Joint Surgery* incorporó en sus normas de publicación estos criterios<sup>52</sup>.

2) Bhandari<sup>53</sup>, del grupo de MBE de Canadá, publica una guía de cómo interpretar y diseñar una revisión sistemática en Cirugía Ortopédica y contribuye con la publicación de un glosario<sup>54</sup>.

3) La celebración del I Symposium sobre diseño, análisis y evaluación crítica de investigación en clínica ortopédica, cuyas ponencias se publicaron en el número de agosto de 2003 de la revista *Clinical Orthopaedics and Related Research*.

## VALORACIÓN DE RESULTADOS EN EL MIEMBRO SUPERIOR

Como ocurre en las demás áreas de nuestra especialidad, la adecuada valoración de los resultados obtenidos con diferentes tratamientos en la patología del miembro superior requiere la utilización de varias herramientas<sup>1,2</sup>. La exploración física y las escalas de puntuación clásicas proporcionan una información extremadamente valiosa. Sin embargo, la valoración de resultados se considera actualmente incompleta si no se valoran otros aspectos, como la intensidad del dolor, la satisfacción y la calidad de vida percibida por el paciente<sup>55</sup>. En este apartado se describen las herramientas más útiles para la valoración de resultados en el miembro superior.

### Dolor y satisfacción

La mayor parte de la patología del miembro superior se manifiesta en forma de dolor. Por ello, una parte fundamental de la valoración de resultados la constituye la valoración del dolor residual que el paciente pueda presentar finalizado un determinado tratamiento o intervención. Asimismo, una

forma sencilla y directa de determinar si el paciente considera que el tratamiento o intervención le ha sido de ayuda es preguntarle sobre su grado de satisfacción global. Lo ideal es valorar tanto el dolor como la satisfacción utilizando escalas analogicovisuales (EAV)<sup>56-58</sup>.

### Exploración física y escalas de valoración clásicas

La exploración física continúa siendo una herramienta indispensable para la valoración de los resultados. Ciertos parámetros como la movilidad articular en diferentes planos permiten una monitorización del cambio a lo largo del tiempo. Esto es especialmente cierto en la extremidad superior, ya que el rango de movilidad del hombro, codo, antebrazo, muñeca y mano presenta una correlación directa con la función de la articulación. Asimismo, el cambio en la respuesta a diferentes maniobras exploratorias ayuda a valorar los resultados del tratamiento; en el miembro superior el mejor ejemplo lo constituye la inestabilidad articular: la desaparición de la inestabilidad con las diferentes maniobras que se emplean en el hombro, el codo y el carpo ayuda a valorar los resultados.

Una primera aproximación a la valoración global del resultado obtenido la constituyen las escalas de puntuación clásicas. En el hombro, las más empleadas son la escala de la sociedad *American Shoulder and Elbow Surgeons* (ASES) y la escala de Constant; la escala de Rowe se emplea selectivamente en casos de inestabilidad<sup>59</sup>. En el codo, la más empleada es la puntuación de la Clínica Mayo (*Mayo Elbow Performance Score* o MEPS). En la muñeca y mano existen diferentes escalas publicadas, como la de Gartland. La mayor parte de éstas valoran tanto el dolor como parámetros medidos por el explorador y capacidad para la realización de diferentes actividades funcionales. Estas escalas continúan siendo válidas, pero no permiten que el paciente valore de forma independiente los resultados del tratamiento.

### Cuestionarios de calidad de vida percibida por el paciente.

Los cuestionarios para la valoración del resultado global (*outcome questionnaires*) se han convertido en herramientas esenciales para la valoración precisa del resultado de nuestros tratamientos. Lo ideal es combinar cuestionarios genéricos y específicos. Dentro de los cuestionarios específicos existen ventajas y desventajas sobre la utilización de cuestionarios que valoren globalmente la extremidad superior o cuestionarios específicos para cada articulación o para determinadas patologías.

### Cuestionarios genéricos

De los diferentes cuestionarios genéricos existentes, probablemente el de mayor utilidad en el campo de la extremidad superior, como en muchos otros, es el SF-36. Este

cuestionario no sólo es el más empleado en multitud de áreas de la Medicina, sino que además está validado en castellano<sup>20,60,61</sup>.

### Cuestionarios específicos

La extremidad superior es una de las áreas en las que más se ha avanzado en el uso de cuestionarios específicos para la valoración de resultados. Como se comenta previamente, existen cuestionarios que intentan valorar toda la extremidad superior globalmente, así como cuestionarios específicos para diferentes áreas anatómicas (hombro, codo, muñeca) o ciertas patologías (por ejemplo, túnel carpiano).

**Cuestionarios globales.** El cuestionario más empleado para la valoración global de la extremidad superior es el cuestionario DASH (*Disabilities of the Arm, Shoulder and Hand*), desarrollado conjuntamente por el *Institute for Work and Health* y la *American Academy of Orthopedic Surgeons* (AAOS)<sup>62</sup>. Este cuestionario ha sido validado en castellano (tabla 7)<sup>63</sup>. El cuestionario consta de 30 preguntas. Además, existen dos módulos opcionales, cada uno de ellos con 4 cuestiones, que se emplean para valorar los síntomas y función de deportistas, artistas y otros trabajadores cuyas demandas funcionales exceden las valoradas por el cuestionario DASH. Puede accederse fácilmente a la información necesaria para la utilización de esta herramienta consultando la página web [www.dash.iwh.on.ca](http://www.dash.iwh.on.ca).

El cálculo de la puntuación utilizando el cuestionario DASH es relativamente complicado. Para poder calcular la puntuación es necesario que se hayan contestado al menos 27 de las 30 cuestiones. La puntuación final se obtiene calculando la media aritmética de las preguntas contestadas, restando 1 y multiplicando por 25: puntuación DASH = ([puntuación cuestiones contestadas/número cuestiones contestadas] - 1) \* 25. Este cálculo proporciona una puntuación que oscila entre 0 y 100; mayor puntuación indica mayor discapacidad. Los módulos adicionales opcionales se calculan de igual forma, no pudiendo calcularse la puntuación si no se han contestado todas las cuestiones de cada uno de estos dos módulos. Se consideran variaciones con trascendencia clínica aquellas que superan los 10 puntos<sup>64</sup>.

Diferentes estudios han utilizado este cuestionario para la valoración de resultados en diferentes patologías. Asimismo, se han realizado estudios para validar el cuestionario en multitud de idiomas<sup>63</sup> y para averiguar los valores que se obtienen aplicando este cuestionario a la población general<sup>65</sup>.

Recientemente se ha desarrollado una versión abreviada del cuestionario DASH que permite una valoración más rápida del resultado<sup>66</sup>. Este cuestionario, denominado *quick-DASH*, consta de 11 preguntas (tabla 8). Existe una elevada correlación entre las puntuaciones de los cuestionarios

*DASH* y *quickDASH*<sup>66</sup>, aunque los autores de estos cuestionarios recomiendan utilizar la versión siempre que sea posible.

**Cuestionarios regionales o por patologías.** Algunos autores han explorado la utilidad de cuestionarios específicos para cada región anatómica (hombro, codo, muñeca, mano) o específicos para determinados tipos de patología. A esta categoría pertenecen multitud de cuestionarios, como el *Shoulder Pain and Disability Index* (SPADI), *Western Ontario Osteoarthritis of the Shoulder Index* (WOOS), *Rotator-Cuff Quality of Life* (RC-QOL), *Oxford Shoulder Scores* (OOS), el *Patient-Rated Wrist-Hand Evaluation* (PRWHE), o el cuestionario *Brigham* del túnel carpiano. Aunque los cuestionarios regionales o por patologías ofrecen una valoración más sensible<sup>67</sup>, la capacidad del cuestionario DASH para detectar cambios parece ser muy similar a la de cuestionarios más específicos<sup>68-71</sup>.

## CONCLUSIONES

Existe un interés creciente por la utilización de diferentes herramientas para la valoración de los resultados obtenidos con las intervenciones o tratamientos de los cirujanos ortopédicos<sup>72</sup>. La exploración física y valoración de parámetros tales como el rango de movilidad articular continúan siendo importantísimos a la hora de valorar los resultados obtenidos. Sin embargo, y especialmente en actividades relacionadas con la investigación, es necesario incorporar otras herramientas para la valoración de resultados, tales como escalas analógicas visuales de valoración del dolor y satisfacción, escalas de puntuación clínica (ASES, Constant, MEPS), y cuestionarios que valoren la calidad de vida percibida por el paciente.

La valoración de los resultados obtenidos en el miembro superior puede realizarse de forma bastante precisa utilizando los cuestionarios *SF-36* y *DASH*. En situaciones en las que el tiempo disponible para rellenar estos cuestionarios sea escaso, la utilización del cuestionario abreviado *quickDASH* proporciona una valoración adecuada de la calidad de vida percibida por el paciente con procesos que afecten el miembro superior. Determinados proyectos de investigación pueden beneficiarse de la utilización de cuestionarios regionales o específicos para ciertas patologías.

## VALORACIÓN DE RESULTADOS EN EL MIEMBRO INFERIOR

### Pelvis y acetábulo

Las lesiones de pelvis y acetábulo son relativamente frecuentes y afectan a grupos de edades de niños y adultos

**Tabla 7.** Versión en castellano del cuestionario DASH, según Rosales<sup>63</sup>

Por favor, puntúe su habilidad o capacidad para realizar las siguientes actividades durante la última semana					
	Ninguna dificultad	Dificultad leve	Dificultad moderada	Mucha dificultad	Imposible de realizar
1. Abrir un bote de cristal nuevo	1	2	3	4	5
2. Escribir	1	2	3	4	5
3. Preparar la comida	1	2	3	4	5
4. Girar la llave para abrir la puerta o arrancar el coche	1	2	3	4	5
5. Empujar y abrir una puerta pesada	1	2	3	4	5
6. Colocar un objeto en estanterías situadas por encima de su cabeza	1	2	3	4	5
7. Realizar tareas duras de la casa como fregar el piso, limpiar paredes y cristales, etc.	1	2	3	4	5
8. Arreglar el jardín o realizar trabajos en el campo	1	2	3	4	5
9. Hacer las camas	1	2	3	4	5
10. Cargar una bolsa del supermercado o llevar un maletín	1	2	3	4	5
11. Cargar con un objeto pesado (más de 5 kilos)	1	2	3	4	5
12. Cambiar una bombilla del techo	1	2	3	4	5
13. Lavarse y secarse el pelo	1	2	3	4	5
14. Lavarse la espalda	1	2	3	4	5
15. Ponerse un jersey o un suéter	1	2	3	4	5
16. Usar un cuchillo para cortar la comida	1	2	3	4	5
17. Actividades de entretenimiento que requieren poco esfuerzo como hacer punto, coser, jugar a las cartas o al dominó	1	2	3	4	5
18. Actividades que requieren algo de esfuerzo para su brazo, hombro o mano, como usar un martillo, jugar al golf, al tenis o a la petanca	1	2	3	4	5
19. Actividades que requieren que su brazo se mueva libremente como nadar	1	2	3	4	5
20. Conducir	1	2	3	4	5
21. Actividad sexual	1	2	3	4	5
22. ¿Durante la última semana, sus problemas del hombro, brazo o mano han interferido con su actividad social habitual con la familia, sus amigos o compañeros de trabajo?	No, para nada	Un poco	Regular	Bastante	Mucho
23. ¿Durante la última semana ha tenido usted dificultad para realizar su trabajo u otras actividades diarias debido a problemas en su hombro, brazo o mano?	No, para nada	Un poco	Regular	Bastante	Mucho
Por favor, ponga puntuación a la gravedad o severidad de los siguientes síntomas					
	Ninguno	Leve	Moderado	Grave	Muy grave
24. Dolor en el hombro, brazo o mano	1	2	3	4	5
25. Dolor en el hombro, brazo o mano cuando realiza cualquier actividad específica	1	2	3	4	5
26. Sensación de calambres, hormigueos o de electricidad en el hombro, brazo o mano	1	2	3	4	5
27. Debilidad o falta de fuerza en el hombro, brazo y mano	1	2	3	4	5
28. Rigidez o falta de movilidad en el hombro, brazo y mano	1	2	3	4	5
29. ¿Durante la última semana, ha tenido dificultad para dormir, debido a dolor en el hombro, brazo o mano?	No	Leve	Moderada	Grave	Dificultad extrema que me impedía dormir
30. Me siento con menos capacidad, confianza y utilidad debido a mi problema con el hombro, brazo o mano	1 Completamente falso	2 Falso	3 No lo sé	4 Cierto	5 Completamente cierto

Tabla 8. Cuestionario quickDASH

<b>QuickDASH</b>					
<b>Please rate your ability to do the following activities in the last week by circling the number below the appropriate response.</b>					
	NO DIFFICULTY	MILD DIFFICULTY	MODERATE DIFFICULTY	SEVERE DIFFICULTY	UNABLE
1. Open a tight or new jar.	1	2	3	4	5
2. Do heavy household chores (e.g., wash walls, floors).	1	2	3	4	5
3. Carry a shopping bag or briefcase.	1	2	3	4	5
4. Wash your back.	1	2	3	4	5
5. Use a knife to cut food.	1	2	3	4	5
6. Recreational activities in which you take some force or impact through your arm, shoulder or hand (e.g., golf, hammering, tennis, etc.).	1	2	3	4	5
	NOT AT ALL	SLIGHTLY	MODERATELY	QUITE A BIT	EXTREMELY
7. During the past week, to what extent has your arm, shoulder or hand problem interfered with your normal social activities with family, friends, neighbours or groups?	1	2	3	4	5
	NOT LIMITED AT ALL	SLIGHTLY LIMITED	MODERATELY LIMITED	VERY LIMITED	UNABLE
8. During the past week, were you limited in your work or other regular daily activities as a result of your arm, shoulder or hand problem?	1	2	3	4	5
<b>Please rate the severity of the following symptoms in the last week. (circle number)</b>					
	NONE	MILD	MODERATE	SEVERE	EXTREME
9. Arm, shoulder or hand pain.	1	2	3	4	5
10. Tingling (pins and needles) in your arm, shoulder or hand.	1	2	3	4	5
	NO DIFFICULTY	MILD DIFFICULTY	MODERATE DIFFICULTY	SEVERE DIFFICULTY	SO MUCH DIFFICULTY THAT I CAN'T SLEEP
11. During the past week, how much difficulty have you had sleeping because of the pain in your arm, shoulder or hand? (circle number)	1	2	3	4	5

QuickDASH DISABILITY/SYMPTOM SCORE =  $\left( \frac{\text{sum of n responses}}{n} - 1 \right) \times 25$ , where n is equal to the number of completed responses.

A QuickDASH score may not be calculated if there is greater than 1 missing item.

jóvenes como resultado de traumatismos de alta energía y a los ancianos y población enferma, con frecuencia por traumatismos de baja energía. Mientras que la pelvis soporta el esqueleto axial y protege las vísceras abdominales y pelvianas, así como estructuras neurovasculares, el acetábulo desempeña un papel fundamental en la bipedestación y la marcha.

La disrupción del anillo pelviano se acompaña de un alto índice de morbilidad y mortalidad asociadas. Por el contrario, las fracturas del acetábulo presentan bajos niveles de mortalidad, pero muy altos niveles de morbilidad debido a la limitación de la movilidad, dolor y artrosis postraumática.

Las medidas de valoración de resultados de las fracturas de pelvis están poco definidas y raramente validadas. No existe consenso internacional y los diversos centros utilizan diferentes valoraciones en base a la clínica y la radiología. Las más frecuentemente utilizadas incluyen la mortalidad, la morbilidad (lesiones vasculares y de esfínteres, lesiones neurológicas y viscerales, disfunciones sexuales, infecciones, dismetrías y algunas complicaciones generales), dolor y niveles de marcha y función (claudicación)<sup>73-82</sup>.

Los métodos de valoración de las fracturas del acetábulo incluyen además sistemas de puntuaciones modificadas de las evaluaciones de las PTC. Estas valoraciones «adaptadas» no acostumbran a estar validadas. El estudio de la literatura revela numerosos sistemas de puntuación de resultados basados en vuelta al trabajo, invalidez, cuestionario Oswestry, función física SF-36, *Iowa Pelvis Score*, *Majeed Outcome Store*<sup>83-85</sup>. Tanto la radiología como la tomografía axial computarizada (TAC) han sido utilizadas como medida de resultados de las fracturas de pelvis<sup>86</sup>. En realidad sólo la TAC ha probado una alta sensibilidad para identificar las pseudoartrosis y medir las consolidaciones defectuosas. Por el momento cada centro usa su propio sistema de análisis haciendo muy difícil la comparación entre grupos<sup>87</sup>.

Los sistemas de valoración de las fracturas del acetábulo incluyen además de cuestionarios específicos otras escalas de puntuaciones modificadas de las evaluaciones de las PTC. Estos cuestionarios «adaptados» no acostumbran a estar validados e incluyen la medida del dolor, capacidad funcional, amplitud de movimiento articular y algunos métodos de medición de la capacidad de marcha. Aún no se utiliza de rutina la TAC y la resonancia magnética (RM) como métodos de valoración sistemáticos.

En los últimos años está incrementando la evidencia de los sistemas de medida basados en los síntomas del paciente, por encima de las puntuaciones obtenidas de las exploraciones clínicas (Merle d'Aubigné, Toronto Hip Score, Harris Hip Score, Iowa Hip Score)<sup>88</sup>. El desarrollo de la *Musculoskeletal Functional Assessment Score* (MFA) ha añadido un sistema de puntuación fiable y validado para pacientes ortopédicos incluyendo los traumatizados. Este cuestionario ha sido recientemente utilizado para valorar los resultados de las fracturas acetabulares por Borelli et al<sup>89</sup>.

## Cadera

Los efectos de los tratamientos para las lesiones de la cadera se pueden basar en un sustancial número de sistemas de valoración de resultados. Los hay subjetivos como el dolor y objetivos como la amplitud de movimientos y radiográficos. Los hay que hacen referencia a los elementos constitutivos de la articulación y otros referidos al estado de salud del paciente. Los hay referidos a la situación próxima a la acción terapéutica y otros se refieren al efecto a largo término. Algunas de estas valoraciones serán improbables a largo término, como los resultados del tratamiento de las fracturas del cuello del fémur en pacientes de elevada edad. Los resultados se verán afectados por la alta mortalidad esperada en esos pacientes a los pocos años de la intervención.

Los cuestionarios más habituales consideran: mortalidad, dolor, movilidad y deformidad, resultados funcionales, escalas de valoración, indicadores generales de salud, valoración radiológica, análisis de supervivencia y análisis coste-beneficio. Existen variaciones sustanciales en la mortalidad en las series publicadas<sup>90,91</sup>, debidas a diversas razones:

- 1) ¿Es una serie consecutiva de pacientes o una serie seleccionada?
- 2) ¿Cómo se han definido los casos incluidos en la serie? Por ejemplo, pacientes entre ciertas edades o con un tipo determinado de lesión.
- 3) ¿Cuántos casos se han perdido en el seguimiento? Es incorrecto asumir que la mortalidad en los casos perdidos es semejante al grupo entero.
- 4) ¿Cuál es el período de tiempo en que se ha considerado la mortalidad? ¿En el período intrahospitalario, tres, 6, 12 meses, varios años después de la cirugía?
- 5) ¿Las muertes estaban relacionadas con la afección de la cadera o con otras causas? Lo más apropiado es buscar una causa específica de muerte, por ejemplo un embolismo pulmonar.

Existen numerosos métodos para describir el dolor de cadera que incluyen variedad de cuestionarios y escalas analógicas visuales del dolor. Uno de los cuestionarios más frecuentemente utilizado es el de Charnley (1972) que tiene 6 grados en la valoración del dolor. En las PTC no cementadas, sin embargo, hay que considerar por separado el dolor del muslo.

La movilidad es un componente de la mayoría de los cuestionarios de valoración de la cadera. Para que sea útil la movilidad de la cadera debe ser comparada con la cadera contralateral normal o con la movilidad previa al tratamiento. Como quiera que en muchos casos las afecciones de cadera sean bilaterales, la referencia se debe tomar con la movilidad preoperatoria. En algunos casos de fracturas, con afectación degenerativa o postraumática de la cadera contralateral, ninguna de esas referencias son útiles. De todos los movimientos de la cadera la flexión es la medida más segura.

De entre los muchos aspectos funcionales que se ven influenciados por las afecciones de la cadera, para las artroplastias coxofemorales se incluyen la habilidad de la marcha (distancia y velocidad), el uso de bastones, el nivel de trabajo de su ocupación o actividad y la habilidad para las actividades cotidianas (subir y bajar escaleras, atarse los cordones de los zapatos, ponerse y quitarse los calcetines, coger objetos del suelo, etc.).

Se ha desarrollado un sistema estandarizado de valoración de resultados tras una fractura de cadera (SAHFE; *Standardised Audit of Hip Fracture in Europe*) que ha demostrado su validez en estudios comparativos entre diferentes centros<sup>92</sup>. Otros cuestionarios funcionales utilizados para fracturas de cadera se basan en actividades de la vida diaria, como el Barthel Index (Mahoney y Barthel 1965) y el de Katz (1963).

Las escalas de valoración de la cadera se basan en criterios clínicos subjetivos (dolor) o objetivos (movilidad), radiológicos o ambos. Uno de los primeros sistemas utilizados fue el de Ferguson y Howarth (1931) popularizado entre nosotros por Merle d'Aubigné y Postel (1954). Este cuestionario, similar al de los hermanos Judet (1952) se basa en tres niveles de dolor, la movilidad de la cadera y habilidad para la marcha de 1 a 6 grados. Las tres cifras obtenidas se suman para determinar la puntuación final. Ejemplos de estas escalas son la IHS (*Iowa Hip Score*), la escala de Harris, la de Charnley, la escala de la Clínica Mayo y la escala WOMAC (*Western Ontario and McMaster Universities*)<sup>93-95</sup>. Al igual que la escala de Harris, la WOMAC se ha constatado fiable, reproducible y válida. Con la intención de resolver algunos de los problemas de los sistemas de medida, en 1990 se propuso unificar estas escalas, lo que fue llevado a cabo por Katz en 1995<sup>96</sup> y adoptada por la SICOT y por la AAOS<sup>97</sup>.

Para las afecciones de cadera se han usado varios sistemas indicadores de salud como el LEM (*Lower Extremity Measure*), el FRS (*Functional Recovery Score*), el SF-36 y el NHP (*Nottingham Health Profile*). El más utilizado de estos métodos es el SF-36 (2.ª versión).

De entre los sistemas de valoración radiológica de la prótesis total de cadera, destacamos el método de DeLee y Charnley (1976) que valora las zonas radiolucidas alrededor del cemento cotiloideo y el método de Gruen (1979) que divide la zona del vástago femoral en tercios tanto en la radiografía anteroposterior como en la lateral. Analiza cronológicamente las fracturas del cemento y las imágenes de radiolucencia entre el vástago y el cemento y entre el cemento y el hueso, así como la aparición de reacciones escleróticas.

Broker (1973) clasificó las osificaciones heterotópicas, en la radiografía AP, en 4 clases.

Con la llegada de la RM Steinberg realizó una modificación de la clasificación clásica de Ficat para las necrosis de la cabeza femoral, ampliamente extendida en la actualidad.

Los criterios de reducción de las fracturas intracapsulares tratadas mediante osteosíntesis también se basan en la radiología (Garden 1961; Parker y Prior 1993). En los casos tratados con artroplastias tipo Moore, Pryor (1990) describió un método de valoración del resultado técnico y para las fracturas pretrocantéreas Parker y Prior en 1993 y Baumgartner y Solberg en 1997 describieron sendos sistemas de valoración del grado de corrección de la fractura y del posicionamiento de la osteosíntesis<sup>98,99</sup>.

Los tres problemas más importantes con los análisis de supervivencia son:

- 1) Los resultados de los pacientes que dejan el seguimiento.
- 2) El análisis de grupos reducidos.
- 3) La definición de fracaso.

Cada año un número de pacientes morirá y otros dejarán el seguimiento. La asunción fundamental de los análisis de supervivencia es que el grupo de retirados tenga el mismo porcentaje de fracasos que los que no abandonan el seguimiento. Es esencial que el número de pacientes perdidos en el seguimiento sean incluidos en el análisis<sup>100,101</sup>. En los análisis de supervivencia es frecuente que sólo un pequeño número de pacientes sea seguido por períodos largos.

Es muy importante definir claramente el concepto de fracaso, por ejemplo de una PTC. Por otra parte, la decisión de recambiar una PTC depende no sólo del estado del paciente, sino también del estado de la lista de espera e incluso de la agresividad del cirujano.

Los métodos de análisis del coste-beneficio más frecuentemente empleados son los QALY (*Quality Adjusted Life Years*). Los QALY y el coste por QALY que se ha determinado para algunos procedimientos ortopédicos como la PTC y para diversos tratamientos de las fracturas de cadera<sup>102</sup>. Hoy día ambos conceptos son aún controvertidos.

## Rodilla

El interés en medir la función de la rodilla empezó por producir sistemas de valoración de resultados en condiciones específicas, como las lesiones ligamentosas, la gonartrosis, las lesiones traumáticas del cartílago, las prótesis y los recambios, entre otras. Los instrumentos clásicos para medir la artroplastia de la rodilla incluían la valoración del dolor, función, alteración de los arcos de movimiento, estabilidad y potencia muscular alrededor de la rodilla. Los rayos X se usaron para estudiar la migración de los componentes, la fractura del cemento o de componentes o las líneas radiolucidas periprotésicas. Los análisis de supervivencia, por otro lado, analizaban los resultados en términos de tiempo de revisión o de fallo<sup>103,105</sup>.

La herramienta ideal debería incluir, sin embargo, un resultado subjetivo del paciente, síntomas, examen clínico, algunos cuestionarios secundarios, al mismo tiempo que re-



sultados de complicaciones, su grado, tiempo de comienzo y duración<sup>100</sup>.

Uno de los primeros cuestionarios que usó estos parámetros fue el del *Hospital for Special Surgery* (HSS). La función era evaluada en relación con la marcha, en escaleras y en plano. Más tarde, se fueron desarrollando otros sistemas de medida de resultados, como el de *New Jersey* y el índice de *Lequesme*.

Konig et al (1989) se dieron cuenta de que, usando el HSS, los resultados de los pacientes empeoraban gradualmente años después de la cirugía de recambio de rodilla, y que ello estaba relacionado con la edad del paciente. La funcionalidad de la rodilla podía permanecer intacta, mientras que el deterioro se producía en el estado de salud general del paciente y en la habilidad en las actividades diarias. Esto estimuló la modificación del HSS hacia el sistema de puntuación de la Sociedad Americana de Rodilla (KSS)<sup>106,107</sup>, que separaba la función de la rodilla de las actividades de los pacientes.

La dificultad se presentó al intentar cuantificar un resultado quirúrgico. Desde el punto de vista del paciente está mejor expresado en términos subjetivos. Los investigadores empezaron a usar cuestionarios de calidad de vida y de salud general como otra medida para valorar los resultados en cirugía. El índice de la WOMAC fue usado satisfactoriamente junto con otros cuestionarios de salud general, incluido el perfil de salud de Nottingham, el SF-36, el SF-12 y la MFA (valoración funcional musculoesquelética)<sup>108-111</sup>.

La escala de actividad diaria de rodilla (ADLS) se desarrolló en base a un análisis de sistemas de medida de resultados existentes incluyendo el Cincinnati, Lysholm, WOMAC y IKDC<sup>112,113</sup>. En el mismo año fue publicado el Cuestionario de Rodilla Oxford<sup>114</sup> y un estudio de Dawson<sup>115</sup> confirmaba su validez y sensibilidad.

Seis de estos cuestionarios han dado un resultado entre 8 y 10: el WOMAC, el Cuestionario Oxford-12, KSS, el índice *Lequesne* de Severidad, el SF-36 y el SF-12. Sólo el KSS (sistema de medida de la sociedad americana de rodilla) incluye medidas objetivas para la validación. Debido a que la mayoría de los clínicos prefieren herramientas subjetivas y autoadministradas, los mejores sistemas de valoración específicos parecen ser el WOMAC y los cuestionarios Oxford-12, mientras que las mejores medidas genéricas serían el SF-36 y el SF-12.

El SF-12 ha dado buenos resultados en la valoración de la salud general en los pacientes con artroplastia de la rodilla, mientras que el cuestionario Oxford-12 ha mostrado el mejor ranking para escalas de enfermedad específica en cuanto a su fiabilidad/exactitud/veracidad, validez de contenido y viabilidad de su uso. Así pues, una combinación de SF-12 y el cuestionario Oxford-12 parece ser la mejor elección.

El primer intento de evaluar los resultados del tratamiento de las lesiones de rodilla fue llevado a cabo por O'

Donoghue (1955). Intentando validar los resultados de reconstrucciones del LCA, ideó una escala de puntuación que incluye un examen objetivo y un cuestionario de 100 preguntas.

Algunos años después, Larson (1974) desarrolló otra escala de 100 puntos basada en categorías subjetivas, objetivas y funcionales. La capacidad funcional se evaluaba andando, corriendo y saltando.

En 1997, Marshall et al desarrollaron una escala de 50 puntos con 4 subcategorías, cada una de las cuales contribuyen al resultado total: parámetros subjetivos (22%), test funcionales (14%), estabilidad de la rodilla (24%) y laxitud de los ligamentos (40%). El cuestionario de rodilla de 100 puntos del *Hospital for Special Surgery* (HSSKS) fue formulado a partir del de Marshall y se introdujeron criterios especiales para actividades deportivas<sup>116</sup>.

La ausencia de un sistema de puntuación estandarizado llevó a la creación de un Comité Internacional de Documentación de la Rodilla (IKDC)<sup>117</sup>. Éste creó una escala uniforme para evaluar los resultados después de una lesión o una reconstrucción del LCA.

El cuestionario *Cincinnati* de rodilla fue originariamente diseñado para lesiones del LCA, pero más tarde sufrió modificaciones para incluir la valoración de todos los ligamentos mayores de la rodilla, así como de lesiones por degeneración meniscal o problemas patelofemorales<sup>118</sup>. Flandry et al (1991) propusieron la primera escala visual analógica para valorar las deficiencias subjetivas de rodilla (el cuestionario Hughston auto-valorado).

El cuestionario de vida diaria de medición de resultados de rodilla (ADLS de KOS) se mostró válido y sensible al cambio, en un estudio para valorar los resultados del tratamiento de rodillas con lesiones del LCA.

La escala deportiva de rodilla de la AAOS fue incluida en el Sistema de Dirección y Evaluación de Datos y Resultados Musculoesqueléticos (MODEMS) para pacientes atletas.

Anderson et al<sup>119</sup> encontraron que para evaluar los resultados de tratamientos de lesiones del LCA habían sido usados 38 sistemas diferentes de puntuación, en 52 artículos publicados durante 10 años en el *American Journal of Sports Medicine* y en el *Journal of Bone and Joint Surgery*. En estudios más recientes<sup>120-123</sup> se identificaron más de 54 medidas de resultados diferentes usadas para valorar la cirugía de LCA, la mayoría de las cuales no habían sido validadas.

Es por eso que se hace muy difícil hacer una propuesta objetiva. Así, parece que, para lesiones del LCA, el cuestionario *Cincinnati* puede ser considerado el más válido, básicamente porque es fácilmente comprensible y cubre todos los aspectos relacionados con los síntomas y funciones de la rodilla. A pesar de la popularidad de la LKSS (*Lysholm Knee Scoring Scale*), el aumento/valoración de sus ítems y el cálculo de un resultado total al sumar dominios no rela-

cionados, disminuyen su valor. La escala ADSL (*Activities of Daily Living Scale*) (una escala básicamente aplicada en pacientes con lesiones del LCA) se ha demostrado como válida. Así, la consideramos también de elección.

Está pendiente de evaluar el KOOS y el ACL-QOL, y el uso de cuestionarios de calidad de vida no está probablemente justificado en los deportistas. El uso tanto de ADSL como de KOOS está recomendado para valorar los resultados de la mayor parte del resto de la cirugía de rodilla.

Para las fracturas de alrededor de la rodilla el único instrumento de valoración específico disponible es el cuestionario de Rasmussen, aunque la valoración subjetiva será suficiente para resultados a largo plazo. No tenemos aún una mejor elección, por lo que recomendamos el uso del WOMAC.

Aunque el principio de usar una herramienta de resultados de rodilla universal para todos los desórdenes de rodilla es atractivo, un sistema así parece no ser práctico debido a las diferencias de demandas de los pacientes de cada grupo.

### Tobillo y pie

Los sistemas de medida de resultados en tobillo y talón comprenden signos, síntomas y complicaciones<sup>124-126</sup> y en los sistemas utilizados habitualmente hay una mezcla de todos ellos.

Entre los signos están la movilidad, la deformidad y la estabilidad, medidas objetivas que están sometidas a significativos errores estando en general no validadas y son difícilmente reproducibles<sup>127,128</sup>. ¿Cuánto de la deformidad y arco de movimiento ha sido afectado por la enfermedad y por el tratamiento? ¿Qué relación exacta tiene la pérdida de función en el resultado final? La comparación de los movimientos del tobillo y de la articulación subastragalina puede compararse con el tobillo contralateral siempre que éste sea normal. Además se ha demostrado que la medición de la movilidad de la articulación subastragalina es difícilmente reproducible y son frecuentes fuentes de error.

La valoración de los síntomas dolor, capacidad funcional y satisfacción cada vez se orientan más hacia cuestionarios centrados en los pacientes. El FFI (*Foot Function Index*) es un sistema de puntuación regional plenamente validado pero se ha utilizado fundamentalmente en artritis reumatoide excluyendo los casos operados y aquellos con deformidades fijas. También se ha utilizado el SF-36 para valorar la respuesta global del paciente a la intervención<sup>129</sup>.

La diversidad en el nivel de satisfacción de los pacientes con estos y otros sistemas de medida puede explicarse por las diferentes expectativas creadas en ellos frente al tratamiento. A pesar de ello, dadas las limitaciones reconocidas en los sistemas de medida objetivos, se ha generalizado el uso de puntuaciones subjetivas. Algunos procedimientos, que pueden resultar beneficiosos a corto y medio plazo pueden en cambio presentar secuelas a largo plazo por lo que

otro factor a tener en cuenta es el momento apropiado de valoración de la medida<sup>130-132</sup>.

En la artrosis de tobillo parece existir una buena correlación entre los resultados clínicos y las alteraciones radiográficas<sup>133</sup>. En la artrosis las dos características principales utilizadas son la presencia de osteofitos y el grado de reducción de la interlínea articular. Ambos están recogidos en las clasificaciones habituales como la de Mazur (1979), aunque la relevancia de los osteofitos se ha cuestionado posteriormente.

Existen dos cuestionarios especialmente reconocidos para valorar las alteraciones del Aquiles, la de Leppilahti (1998) diseñada para medir las rupturas del tendón operadas y en la que se recogían 7 criterios por igual: dolor, rigidez, pérdida de fuerza en la pantorrilla, restricciones en el calzado, satisfacción del paciente, diferencia en los arcos de movimiento entre los tobillos y la fuerza isocinética del tríceps. Maffulli (1999) señaló que la cuantificación de la función muscular utilizando medidas isocinéticas es una técnica cara, difícil y prolongada y por ello no generalizable.

El segundo sistema de puntuación, el VISA-A (*Victorian Institute of Sports Assessment-Achilles*)<sup>134</sup> que valora 8 parámetros otorgando a cada uno 10 puntos en una escala analógica visual, está diseñado para valorar la severidad de la tendinopatía del Aquiles, ha sido validado y su fiabilidad demostrada en términos de discernir entre tendones normales y los diversos grados de tendinopatía.

Un cuestionario fiable y validado para las lesiones del LLE es el Kaikkonen (1994) que da un máximo de 100 puntos, 40 para síntomas subjetivos y valoración de la función y 60 para 4 test objetivos de la función (uno para el balance, uno para la estabilidad y dos para la fuerza muscular), para la valoración de la laxitud AP y la amplitud de movimiento. Otro sistema de puntuación utilizado es el Karlsson y Peterson (1991), en el que 45 de los 100 puntos son totalmente subjetivos y se refieren a la inestabilidad y dolor. Existe una versión simplificada de Nimon (2001) que tiene 4 categorías de dolor y 4 de inestabilidad subjetiva.

La dificultad de la heterogeneidad de estos sistemas de medida se ha puesto de manifiesto en el metaanálisis realizado por Pijnenburg<sup>135</sup> en el que se demuestra que en el tratamiento de las lesiones del LLE sólo se pudieron comparar tres medidas de resultados, el tiempo de vuelta al trabajo, el dolor residual y la amplitud del paso.

Takakura (1990) describió un sistema de puntuación para valorar las prótesis totales de tobillo, aunque luego lo han utilizado en artrodesis. De nuevo de 100 puntos, 40 eran para el dolor, 20 para la amplitud de movimiento y 40 para la función (incluyendo la habilidad para sentarse al estilo japonés)

La *American Foot and Ankle Society* adoptó el cuestionario de Kitaoka (1994) que ha sido aceptado en muchos de los trabajos de investigación de América y Europa. Permanece no validado para muchas de las indicaciones para las

que fue postulado (fracturas de tobillo y calcáneo, inestabilidad de tobillo, artrodesis y artroplastias)<sup>136</sup>. Rowan, en el 2001, desarrolló un cuestionario para la medición del dolor crónico en el pie<sup>137</sup>.

Para la artrodesis subastragalina y para la triple artrodesis el sistema de puntuación más comúnmente utilizado es el AOFAS adaptado<sup>138,139</sup>. Para la artrodesis tibiocalcánea, Kitaoka y Patzer (1998) adaptaron un sistema de puntuación estrictamente subjetivo.

Uno de los pocos cuestionarios específicos a las fracturas de tobillo es el de Olerud y Molander (1984). Se trata de un sistema totalmente subjetivo de 100 puntos, 25 para el dolor, 20 para la rigidez y 55 para la función. Entre los sistemas no específicos el más utilizado es el SF-36. De entre los cuestionarios específicos comúnmente usados en las fracturas de calcáneo, el más aceptado es el *Maryland Foot Score* con 45 puntos para el dolor y 55 para la función. En los últimos años se ha producido un aumento del interés sobre los sistemas de medida en el tobillo, centrados en el paciente y subjetivos. Parece apreciarse una orientación clara hacia las valoraciones subjetivas por encima de las medidas objetivas. Para el tobillo el VISA-A y en su defecto el SF-36 son los considerados patrón oro, a pesar de sus claras limitaciones.

## VALORACIÓN DE RESULTADOS EN LA COLUMNA VERTEBRAL

Tradicionalmente el éxito en el tratamiento de los trastornos de la columna vertebral se mide en términos de mortalidad, cambios fisiológicos (por ejemplo: conducción nerviosa), o mejoría en la exploración física (debilidad muscular). Más recientemente se ha introducido en la evaluación de resultados considerar los cuestionarios rellenos por los pacientes, su función física y su estado de salud. Este tipo de evaluación de resultados es relativamente joven, pero ha crecido en el número y tipo de mediciones de forma exponencial. El problema en la actualidad es cómo un cirujano del raquis muy ocupado puede escoger entre estas medidas, o bien de otra forma, cómo un investigador en la patología de la columna puede seleccionar un instrumento, fiable, válido y coste-efectivo para usar en proyectos de investigación.

Es por ello que debemos plantearnos determinadas preguntas clave sobre la medición de resultados a tener en cuenta en patología vertebral. Algunas áreas son todavía controvertidas como qué conceptos de satisfacción debemos usar o cómo medir la reducción de la productividad en el trabajo. No obstante, existe un consenso suficiente en la evaluación de la función y del estado de salud del que disponen en forma de guía los investigadores. Estas guías están reflejadas en varias revisiones publicadas en los últimos años<sup>140-146</sup> o en guías donde se establece cómo medir los estados de salud<sup>147-150</sup>.

En primer lugar nos debemos plantear qué resultados queremos medir o evaluar. Para contestar esta pregunta nos referimos al modelo conceptual de Wilson y Cleary (1995). Este modelo especifica 5 tipos de resultados. En la figura 1 se establece la relación entre ellos. Los dos primeros factores, biológicos, fisiológicos y síntomas son variables tradicionalmente clínicas, que se encuentran en la mayoría de estudios y suelen consistir en datos del examen físico y pruebas de laboratorio. Los tres siguientes son medidas basadas en la opinión del paciente:

1) Las medidas de función se basan en la posibilidad de realizar determinadas tareas que usualmente cubren los dominios físico, social y psicológico.

2) La percepción general de salud engloba varios aspectos reflejados en valores subjetivos del paciente.

3) La calidad global de salud es una medida general del bienestar del paciente.

## Función específica del raquis

¿Qué medida concreta debemos utilizar para valorar la función? Los cuestionarios específicos más utilizados son el de Cuestionario de discapacidad de Roland-Morris (RDQ) y el índice de discapacidad de *Oswestry* (ODI). El RDQ mide la limitación en 24 actividades ocasionadas por el dolor de espalda mientras que el ODI valora 10 apartados sobre la influencia del dolor en la función. Sin embargo, en la práctica ambos instrumentos son similares, fáciles de usar, fiables y válidos. Para los rangos más extremos que muestran efecto cielo y suelo (*ceiling and floor effect*), el ODI puede ser una mejor opción, en especial en poblaciones con niveles de gran discapacidad, mientras que el RDQ se utiliza en grupos generales de población en que los valores de discapacidad están en la franja de valores más bajo.

## Estado general de salud

De los cuestionarios de salud el SF-36 muestra el mejor balance entre la dimensión del cuestionario, la fiabilidad, la validez, el índice de respuestas en poblaciones afectas de dolor raquídeo. El SF-36 tiene 8 escalas o dimensiones: función física, dolor corporal, papel físico, salud general, vitalidad, función social, mental y emocional, que pueden sumarse en dos medidas: mental y física. El SF-36 tiene la ventaja de conocerse los valores de grandes grupos de población normal y de diversos grupos de población incluyendo también individuos con dolor de columna, y además es el cuestionario con mayor número de traducciones y validaciones internacionales.

Existen dos versiones, la aguda (una semana) y la crónica (más de un mes), de manera que en el dolor agudo que cambia en pocos días debe aplicarse el primer tipo y en el dolor crónico el segundo. Existe una forma simplificada el SF-12 de uso en aquellos casos de seguimiento de gran número de pacientes sometidos a múltiples cuestionarios.

## Dolor

Las medidas de la gravedad del dolor difieren de las medidas de la afección dolorosa. Por ello es recomendable el uso de diferentes cuestionarios según qué particularidad del dolor queramos evaluar. La gravedad del dolor debe evaluarse mediante breves cuestionarios como la escala analógica visual (EAV) del dolor, y en cambio para la afección dolorosa es recomendable usar las dimensiones del dolor de escalas más amplias como el SF-36. En estos casos es importante medir el dolor no en un momento puntual, sino en el transcurso del tiempo. La escala del dolor corporal del SF-36 mide la intensidad del dolor (en 6 niveles: ninguno, muy poco, un poco, moderado, mucho, muchísimo) y su interferencia con las actividades (en 5 niveles: nada, un poco, regular, bastante y mucho).

Cuando precisamos atribuir el dolor a una condición específica es aconsejable usar la escala gradual de dolor crónico (EGDC), que incluye además preguntas sobre el intervalo de dolor (el momento de peor dolor, el promedio de dolor o un intervalo concreto de dolor).

## Discapacidad laboral

Los estados de salud relacionados con el trabajo han merecido poco interés en las publicaciones. Como mínimo el estatus laboral debe ser medido en la primera y en la última visita de un seguimiento. Una clasificación útil debe incluir el empleo en su trabajo habitual, asignación de un trabajo más ligero, baja laboral, indemnización por baja laboral o no, jubilación por enfermedad o por otras razones, estudiante, ama de casa, jubilado discapacitado. Es también recomendable medir el tiempo de trabajo perdido y el momento de reincorporación laboral. El SF-36 tiene preguntas sobre la dimensión de las limitaciones funcionales que incluyen el trabajo, pero no de forma específica, siendo más útil para esta dimensión específica el cuestionario de limitación laboral WL-26.

## Satisfacción subjetiva

Es una dimensión importante en la evaluación de resultados, de la que existen diferentes aproximaciones. Puede solicitarse la satisfacción por el trato o bien por el resultado del tratamiento específicamente médico. Las medidas pueden ser para la población general o para determinados grupos.

Existe cierta ambigüedad en la evaluación de la satisfacción y para el problema específico del dolor en la columna se han descrito diversas medidas. La escala de satisfacción del paciente PSS es una escala multivariante con 17 preguntas, que evalúan la información recibida, el soporte emocional y la efectividad de un tratamiento prescrito. Cada apartado tiene 5 categorías de respuesta:

1 Totalmente de acuerdo a 5 totalmente en desacuerdo. El PSS mide fundamentalmente la satisfacción en el trato y

siendo recomendable añadir un apartado de sugerencias libres por parte del paciente al final del cuestionario.

En resumen para la medición de la función específica en la columna vertebral, los cuestionarios de Roland y Morris y el índice de discapacidad de Oswestry son los más recomendables. El SF-36 es el más útil para evaluar el estado de salud global y el dolor crónico. Para el estatus laboral puede utilizarse la dimensión funcional del SF-36 o escalas más específicas como el WL-26 y finalmente el PSS para la satisfacción del paciente.

## BIBLIOGRAFÍA

1. Sánchez-Sotelo J. Instrumentos de valoración de la salud y su aplicación a la evaluación de resultados en Cirugía Ortopédica y Traumatología. En Actualizaciones en Cirugía Ortopédica y Traumatología. Actualizaciones SECOT 2. Barcelona: Masson SA; 2001. p. 3-14.
2. Sánchez-Sotelo J. Instrumentos de valoración del estado de salud en Traumatología y Cirugía Ortopédica. Rev Ortop Traumatol. 2001;48:304-14.
3. Vitale MG, Keith MW. The importance of outcomes research. AAOS Online Bulletin. Disponible en: [www.aaos.org/wordhtml/bulletin/dec03/fline2.htm](http://www.aaos.org/wordhtml/bulletin/dec03/fline2.htm) December 2003.
4. Donabedian A. La calidad de la asistencia. JANO. 1989; 36:103-10.
5. Amadio PC. Outcome assessment in hand surgery. Clin Plast Surg. 1997;24:191-4.
6. Hernández Vaquero D, Barrera Cadenas JL. Sistemas de evaluación de los resultados en las artroplastias. Rev Ortop Traumatol. 1999;4:245-51.
7. Johanson NA, Liang MH, Daltroy L, Rudicel S, Richmod J. American academy of orthopaedic surgeons lower limb outcomes assessment instruments. Reliability, validity and sensitivity to change. J Bone Joint Surg Am. 2004;86A:902-9.
8. Johanson NA. Outcomes assessment. En: Callaghan JJ, Rosenberg AG, Rubash HE, editors. The Adult Hip. Philadelphia: Lippincott-Raven Publishers; 1998. p. 853-63.
9. Keller RB. Measuring outcomes. J Orthop Res. 1996;14: 171-2.
10. Keller RB. Outcomes research in orthopaedics. J Am Acad Orthop Surg. 1993;1:122-9.
11. Relman AS. Assessment and accountability: the third revolution in medical care. N Engl J Med. 1988;319:1220-2.
12. Swiontkowski MF, Buckwalter JA, Keller RB, Haralson R. The outcomes movement in orthopedic surgery: where we are and where we should go. J Bone Joint Surg Am. 1999; 81A:732-40.
13. Swiontkowski ME, Engelberg R, Martin DP, Agel J. Short musculoskeletal function assessment questionnaire: validity, reliability, and responsiveness. J Bone Joint Surg Am. 1999; 81A:1245-60.
14. Keller RB, Rudicel SA, Liang MH. Outcomes research in orthopaedics. AAOS Instr Course Lect. 1994;43:599-611.
15. Gartland JJ. Orthopaedic clinical research. Deficiencies in experimental design and determinations of outcome. J Bone Joint Surg Am. 1988;70A:1357-64.
16. Kirschner B, Guyatt G. A methodological framework for assessing health indices. J Chron Dis. 1985;38:27-36.
17. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. Ann Intern Med. 1993;118:622-9.

18. Gross M. A critique in the methodologies used in clinical studies of hip-joint arthroplasty published in the english-language orthopedic literature. *J Bone Joint Surg Am.* 1988; 70A:1364-71.
19. Kantz ME, Harris WJ, Levitsky K, Ware JR, Davis AR. Methods for assessing condition-specific and generic functional status outcomes after total knee replacement. *Med Care.* 1992;30 5 Supl: MS240-52.
20. Alonso J, Regidor E, Barrio G, Prieto L, Rodríguez C, De la Fuente L. Valores poblacionales de referencia de la versión española del cuestionario de salud SF-36. *Med Clin (Barc).* 1998;111:410-6.
21. Badía X, Salamero M, Alonso J. La medida de la salud. Guía de escalas de medición en español. Ed. Lilly; 1999. p. 349.
22. Keller RB, Ware JE, Bentler PM, Aaronson NK, Alonso J, Apolone G, et al. Use of structural equation modelling to test the construct validity of the SF-36 health survey in ten countries: results from the IQOLA project. *International Quality of Life Assessment. J Clin Epidemiol.* 1998;51: 1179-88.
23. Keller RB, Soule DN, Wennberg JE, Hanley DF. Dealing with geographic variations in the use of hospitals. The experience of the Maine Medical Assessment Foundation Orthopaedic Study Group. *J Bone Joint Surg Am.* 1990;72A: 1286-93.
24. Martin DP, Engelberg R, Agel J, Snapp D, Swiontkowski ME. Development of musculoskeletal extremity health assessment instrument: the Musculoskeletal Functional Assessment Instrument. *J Orthop Res.* 1996;14:173-81.
25. McHorney CA, Ware JE, Raczek AE. The MOS 36-item Short-Form Health Survey (SF-36): 11. Psychometric and clinical test of validity in measuring physical and mental health constructs. *Med Care.* 1993;3:247-63.
26. AETS. Índices y escalas utilizados en ciertas tecnologías de la prestación ortoprotésica (Protetización del Sistema Osteoarticular). Instituto de Salud Carlos III. Madrid: Ministerio de Sanidad y Consumo; 2002.
27. Jovell AJ, Navarro-Rubio MD. Evaluación de la evidencia científica. *Med Clin (Barc)* 1995;105:740-43.
28. Weed DL. Methodologic implications of the Precautionary Principle: causal criteria. *Int J Occup Med Environ Health.* 2004;17(1):77-81.
29. Sotomayor H. Revistas médicas en castellano. *RMS Revista Médica de Santiago.* 2000; 3(16) (Chile). Disponible en: ([http://www.rms.cl/Numeros\\_anteriores/rms\\_16/carta\\_editor.htm](http://www.rms.cl/Numeros_anteriores/rms_16/carta_editor.htm)).
30. Hersh WR, Hickam DH. How well do physicians use electronic information retrieval systems? *JAMA.* 1998;280(15): 1347-52.
31. Miser WF. Critical Appraisal of the Literature. *J Am Board Fam Pract.* 1999;12(4):315-33.
32. Pritchard P. Compact disk for GPs-this year-next year? *J Informat Prim Care (On line).* 1995;September.
33. Sackett, et al. Evidence-based medicine. A new approach to teaching the practice of medicine. Evidence-Based Medicine Working Group. *JAMA.* 1992;268(17):2420-5.
34. García F. Un cambio de paradigma en la práctica de medicina Medicina Basada en la Evidencia. Publ. Merck & Co, Inc.; 1998.p. 9-13.
35. Hubbs PR, Rindfleisch TC, Godin P, Melmon KL. Medical information on the Internet. *JAMA.* 1998;280(15):1363.
36. Uriarte J. Medicina Basada en la Evidencia Guías de Práctica Clínica. Curso Lectura Crítica Bibliografía. Hospital de Basurto. 30 noviembre, 12 y 13 de diciembre 2000.
37. Tutorial de Medline en castellano. Disponible en: [http://www.fisterra.com/recursos\\_web/no\\_explor/pubmed.asp](http://www.fisterra.com/recursos_web/no_explor/pubmed.asp).
38. Bachmann LM, Coray R, Estermann P, Ter Riet G. Identifying Diagnostic Studies in MEDLINE: Reducing the Number Needed to Read. *J Am Med Informatics Association.* 2002;9:653-8.
39. Chow TKF, To E, Colin S, Goodchild CS, McNeil JJ. A Simple, Fast, Easy Method to Identify the Evidence Base in Pain-Relief Research: Validation of a Computer Search Strategy Used Alone to Identify Quality Randomized Controlled Trials. *Anesth Analg.* 2004;98:1557-65.
40. Devillé WLJM, Bezemer PD, Bouter LM. Publications on diagnostic test evaluation in family medicine journals: an optimal search strategy. *J Clin Epidemiol.* 2000;53:65-9.
41. Forrest JL, Miller SA. Evidence-Based Decision Making in Action: Part 1 - Finding the Best Clinical Evidence. *J Contemp Dent Pract.* 2002;3:10-26.
42. Bickley SR, Harrison JE. How to find the evidence. *J Orthodontics.* 2003;30:72-8.
43. Bhandari M, Morrow F, Kulkarni A, Tornetta P 3rd. Meta-analyses in orthopaedic surgery. A systematic review of their methodologies. *J Bone Joint Surg Am.* 2001;83A: 15-24.
44. Haynes RB, McKibbin KA, Wilczynski NL, Walter SD, Werre SR. Hedges Team. Optimal search strategies for retrieving scientifically strong studies of treatment from Medline: analytical survey. *BMJ.* 2005;330(7501): 1179-82.
45. Haynes RB, Wilczynski N, McKibbin KA, et al. Developing optimal search strategies for sound studies detecting clinically in MEDLINE. *J Am Med Inform Assoc.* 1994; 1:447-58.
46. Allison JJ, Kiefe CI, Weissinan NW, et al. The art and science of searching MEDLINE to answer clinical questions. Finding the right number of articles. *Int J Technol Assess Health Care.* 1999;15:281-96.
47. Rudicel S, Esdaile J. The randomized clinical trial in orthopaedics: obligation or option? *J Bone Joint Surg Am.* 1985; 67A(8):1284-93.
48. Kiter E, Karatosun V, Günel I. Do orthopaedic journals provide high-quality evidence for clinical practice? *Arch Orthop Trauma Surg.* 2003;123(2-3):82-5.
49. Zanolli G, Romanini E, Padua R, Traina GC, Massari L.. EBM in musculoskeletal diseases: where are we? *Acta Orthop Scand.* 2002; Supl 305:73-4-7.
50. Beaton DE, Schemitsch E. Measures of health-related quality of life and physical function. *Clin Orthop.* 2003; 413:90-105.
51. Bhandari M, Montori VM, Devereaux PJ, Wilczynski NL, Morgan D, Haynes RB. Doubling the Impact: Publication of Systematic Review Articles in Orthopaedic Journals. *J Bone Joint Surg Am.* 2004;86A:1012-6.
52. Wright JG, Swiontkowski MF. Introducing a new Journal section: Evidence-Based Orthopaedics. *J Bone Joint Surg Am.* 2000;82A(6):759-60.
53. Bhandari M, Guyatt GH, Montori V, Devereaux PJ, Swiontkowski MF. User's Guide to the Orthopaedic Literature: How to Use a Systematic Literature Review *J Bone Joint Surg Am.* 2002;84A:1672-82
54. Bhandari M, Tornetta P3rd, Guyatt GH. Glossary of Evidence-Based Orthopaedic Terminology. *Clinical Orthopaedics and Related Research* 2003; August: 413,158-63.
55. Harvie P, Pollard TC, Chennagiri RJ, Carr AJ. The use of outcome scores in surgery of the shoulder. *J Bone Joint Surg Br.* 2005;87B:151-4.

56. Angst F, Aeschlimann A, Michel BA, Stucki G. Minimal clinically important rehabilitation effects in patients with osteoarthritis of the lower extremities. *J Rheumatol.* 2002;29:131-8.
57. Bijur PE, Silver W, Gallagher EJ. Reliability of the visual analog scale for measurement of acute pain. *Acad Emerg Med.* 2001;8:1153-7.
58. Salaffi F, Stancati A, Silvestri CA, Ciapetti A, Grassi W. Minimal clinically important changes in chronic musculoskeletal pain intensity measured on a numerical rating scale. *Eur J Pain.* 2004;8:283-91.
59. Kirkley A, Griffin S, Dainty K. Scoring systems for the functional assessment of the shoulder. *Arthroscopy.* 2003;19:1109-20.
60. Alonso J, Lamarca R, Martí-Valls J. The pain and function of the hip (PFH) scale: a patient-based instrument for incasuring outcome after total hip replacement. *Orthopedics.* 2000;23:1273-7.
61. 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.
62. Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand) [corrected]. The Upper Extremity Collaborative Group (UECG). *Am J Ind Med.* 1996;29:602-8.
63. Rosales RS, Delgado EB, Diez de la Lastra-Bosch I. Evaluation of the Spanish version of the DASH and carpal tunnel syndrome health-related quality-of-life instruments: cross-cultural adaptation process and reliability. *J Hand Surg [Am].* 2002;27:334-43.
64. Gummesson C, Atroshi I, Ekdahl C. The disabilities of the arm, shoulder and hand (DASH) outcome questionnaire: longitudinal construct validity and measuring self-rated health change after surgery. *BMC Musculoskelet Disord.* 2003;4:11.
65. Hunsaker FG, Cioffi DA, Amadio PC, Wright JG, Caughlin B. The American Academy of Orthopaedic Surgeons Outcomes Instruments: Normative Values from the General Population. *J Bone Joint Surg Am.* 2002;84A:208-15.
66. Beaton DE, Wright JG, Katz JN. Development of the Quick-DASH: Comparison of Three Item-Reduction Approaches. *J Bone Joint Surg Am.* 2005;87A:1038-46.
67. Gay RE, Amadio PC, Johnson JC. Comparative responsiveness of the disabilities of the arm, shoulder, and hand, the carpal tunnel questionnaire, and the SF-36 to clinical change after carpal tunnel release. *J Hand Surg Am.* 2003;28:250-4.
68. Beaton DE, Katz JN, Fossel AH, Wright JG, Tarasuk V, Bombardier C. Measuring the whole or the parts? Validity, reliability, and responsiveness of the Disabilities of the Arm, Shoulder and Hand outcome measure in different regions of the upper extremity. *J Hand Ther.* 2001;14:128-46.
69. Davis AM, et al. Measuring disability of the upper extremity: a rationale supporting the use of a regional outcome measure. *J Hand Ther.* 1999;12:269-74.
70. Greenslade JR, Mehta RL, Belward P, Warwick DJ. Dash and Boston questionnaire assessment of carpal tunnel syndrome outcome: what is the responsiveness of an outcome questionnaire? *J Hand Surg Br.* 2004;29B:159-64.
71. MacDermid JC, Tottenham V. Responsiveness of the disability of the arm, shoulder, and hand (DASH) and patient-rated wrist/hand evaluation (PRWHE) in evaluating change after hand therapy. *J Hand Ther.* 2004;17:18-23.
72. 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; discussion 735-7.
73. Biffl W, Smith WR, Moore EE, González RJ. Evolution of a multidisciplinary clinical pathway for the management of unstable patients with pelvic fractures. *Ann Surg.* 2001;233:843-50.
74. Brenneman ED, Katyal D, Boulanger BR. Long term outcomes in open pelvic fractures. *J Trauma.* 1997;42:773-7.
75. Cole JD, Blum DA, Ansel LJ. Outcome after fixation of unstable posterior pelvic ring injuries. *Clin Orthop.* 1996;329:160-79.
76. Leutenegger A, Ruedi T. Fractures of the acetabulum and pelvic ring: epidemiology and clinical outcome. *Swiss Surg.* 1999;5:47-54.
77. Liebergall M, Low J, Goldvirt M, Matan Y, Segal D. Acetabular fractures. Clinical outcome of surgical treatment. *Clin Orthop.* 1999;366:205-16.
78. Oliver C, Twaddle B, Agel J. Outcome after pelvic ring fractures: evaluation using the medical outcomes short form SF-36. *Injury.* 1996;27:635-41.
79. Rice JKM, Dolan M, Cox M, Khan H, McElwain J. Comparison between clinical and radiologic outcome measures after reconstruction of acetabular fractures. *J Orthop Trauma.* 2002;16:82-6.
80. Tornetta P, Matta J. Outcome of operatively treated unstable posterior pelvic ring disruptions. *Clin Orthop.* 1996;329:186-93.
81. Tornetta P, Dickson K, Matta M. Outcome of rotationally unstable pelvic ring injuries treated operatively. *Clin Orthop.* 1996;329:147-51.
82. Van den Bosch EW, Van der Kleyn R, Hoegervorst M. Functional outcome of internal fixation for pelvic ring fractures. *J Traum.* 1999;47:365-71.
83. American Academy of Orthopaedic Surgeons: Joint Motion: Method of measuring and recording. Chicago: American Academy of Orthopaedic Surgeons; 1965.
84. American Academy of Orthopaedic Surgeons: Scoring Algorithms for the Lower Limb Outcomes Data Collection Instrument, Version 2.0. Chicago: American Academy of Orthopaedic Surgeons; 1988.
85. Deo STS, Pandey R, El-Saied G, Willett K, Woriocok P. Operative treatment of acetabular fractures in Oxford. *Injury.* 2001;32: 581-6.
86. Miranda M, Riemer B, Butterfield M, Burke C. Pelvic ring injuries: a long term functional outcome study. *Clin Orthop.* 1996;329:152-9.
87. Pynsent P, Fairbank J, Carr A, editors. Outcome Measures in Orthopaedics and Orthopaedic Trauma. London: Arnold; 2004.
88. Soderman P, Malchau H. Is the Harris hip score system useful to study the outcome of total hip replacement? *Clin Orthop.* 2001;384:189-97.
89. Borrelli JGC, Ricci W, Wigner I. Functional outcome after isolated acetabular fractures. *F Orthop Trauma.* 2002;16:73-81.
90. Parker MJ, Palmer CR. A new mobility score for predicting mortality after hip fracture. *J Bone Joint Surg Br.* 1993;75B:797-8.
91. Parker MJ, Currie CT, Mountain JA. Standardised Audit of Hip Fracture in Europe (SAHFE). *Hip International.* 1998;8(10):15.
92. Jagal S, Lakhani Z, Schatzer J. Reliability, validity, and responsiveness of the lower extremity measure for patients

- with a hip fracture. *J Bone Joint Surg Am.* 2000;82A:955-62.
93. Engelberg R, Martin DP, Agel J. Musculoskeletal function assessment instrument: criterion and construct validity. *J Orthop Res.* 1996;14:182-92.
  94. Fitzpatrick R, Morris R, Hajat S. Oxford Hip Score: the value of short and simple measures to assess outcomes for patients of total hip replacement surgery. *Qual Health Care.* 2000;9:146-50.
  95. Guyton PG. Theoretical limitations of the AOFAS scoring systems: an analysis using Monte Carlo modelling. *Foot Ankle Int.* 2001;22:779-87.
  96. Katz JN, Phillips CB, Poss R, Harrast JJ. The validity and reliability of a total hip arthroplasty outcome evaluation questionnaire. *J Bone Joint Surg Am.* 1995;77A:1528-34.
  97. Liang MH, Katz JN, Phillips C. The total hip arthroplasty outcome evaluation form of the American Academy of Orthopaedic Surgeons. *J Bone Joint Surg Am.* 1991;73A:639-46.
  98. Zuckerman JD, Koval KJ, Aharonoff. A functional recovery score for elderly hip fracture patients J. *Development. J Orthop Trauma.* 2000;4:20-5.
  99. Zuckerman JD, Koval KJ, Aharonoff GB. A functional recovery score for elderly hip fracture patients. Validity and reliability. *J Orthop Trauma.* 2000;14:26-30.
  100. Carr AJ, Morris RW, Murray DW, Pynsent PB. Survival analysis in joint replacement surgery. *J Bone Joint Surg Br.* 1993;75B:178-82.
  101. Murray DW, Carr AJ, Bulstrode C. Survival analysis of joint replacements. *J Bone Joint Surg Br.* 1993;75B:697-704.
  102. Parker MJ, Myles JW, Anand JK, Drewett R. Cost-benefit analysis of hip fracture treatment. *J Bone Joint Surg.* 1992;74B:261-4.
  103. Dorey E, Amstutz HC. The validity of survivorship analysis in total joint arthroplasty. *J Bone Joint Surg Am.* 1989;71A:544-8.
  104. Dunbar MJ, Robertsson O, Ryd L, Lidgren L. Appropriate questionnaires for knee arthroplasty. *J Bone Joint Surg Br.* 2001;83B:339-44.
  105. Dunbar MJ, Robertsson O, Ryd L, Lidgren L. Translation and validation of the Oxford-12 item knee score for use in Sweden. *Acta Orthop Scand.* 2000;71:268-74.
  106. Insall JN, Dorr LD, Scott RD, Scott WN. Rationale of the Knee Society clinical rating system. *Clin Orthop.* 1989;248:13-4.
  107. Hóher J, Bach T, Münster A, Bouillon B, Tiling T. Does the mode of data collection change results in a subjective knee score? *Am J Sports Med.* 1997;25:642-7.
  108. Lingard EA, Katz JN, Wright EA, Wright RJ, Sledge CB. Validity and responsiveness of the Knee Society Clinical Rating System in comparison with the SF-36 and WOMAC. *J Bone Joint Surg Am.* 2001;83A:1856-64.
  109. Liow RYL, Walker K, Wajid MA, Bedi G. The reliability of the American Knee Society score. *Acta Orthop Scand.* 2000;71:603-8.
  110. Marx RG, Jones EC, Allen AA. Reliability, validity and responsiveness of four knee outcome scales for athletic patients. *J Bone Joint Surg Am.* 2001;83A:1459-69.
  111. McConnell S, Kolopack P, Davis AM. The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC): a review of its utility and measurement properties. *Arthritis Rheum.* 2001;45:453-61.
  112. Bollen S, Seedhom B. A comparison of the Lysholm and Cincinnati Knee scoring questionnaires. *Am J Sports Med.* 1991;19:189-90.
  113. Risberg MA, Holm I, Steen H, Beynnon BD. Sensitivity to change over time for the IKDC form, The Lysholm score and the Cincinnati knee score. *Knee Surg Sports Traumatol Arthrosc.* 1999;7:152-9.
  114. Harcourt WGV, White SH, Jones P. Specificity of the Oxford knee status questionnaire. *J Bone Joint Surg Br.* 2001;83B:345-7.
  115. Dawson J, Fitzpatrick R, Murray D, Carr A. Questionnaire on the perception of patients about total knee replacement. *J Bone Joint Surg Br.* 1998;80B:63-9.
  116. Barber-Westin SD, Noyes ER, McCloskey JW. Rigorous statistical reliability, validity, and responsiveness testing of the Cincinnati knee rating system in 350 subjects with uninjured, injured, or anterior cruciate ligament-reconstructed knees. *Am J Sports Med.* 1999;27:402-16.
  117. Hefti E, Muller W, Jakob RP, Staubli HU. Evaluation of knee ligament injuries with the IKDC form. *Knee Surg Sports Traumatol Arthrosc.* 1993;1:226-34.
  118. Visentini P, Khan K, Cook J, et al. The VISA score an index of the severity of jumper's knee (patellar tendinosis). *J Sci Med Sport.* 1998;1:22-8.
  119. Anderson AF, Lipscomb AB. Clinical diagnosis of meniscal tears. Description of new manipulative test. *Am J Sports Med.* 1986;14:291-3.
  120. Johnson DS, Smith RB. Outcome measurement of ACL deficient knee - what's the score? *Knee.* 2001;8:51-7.
  121. Muller W, Biedert R, Hefti E, Jakob RP, Munzinger U, Staubli HU. OAK knee evaluation. A new way to assess knee ligament injuries. *Clin Orthop.* 1988;232:37-50.
  122. Sgaglione NA, Del Pizzo W, Fox JM, Friedman MJ. Critical analysis of knee ligament rating scales. *Am J Sports Med.* 1995;23:660-7.
  123. Shapiro ET, Richmond JC, Rockett. The use of generic, patient-based health assessment (SF-36) for evaluation of patients with anterior cruciate ligament injuries. *Am J Sports Med.* 1996;24:196-200.
  124. Heffernan G, Khan E, Awan N, O'Riordain. A comparison of outcomes scores in os calcis fractures. *Irish J Med Sci.* 2000;169:127-8.
  125. Pynsent PB. Choosing an outcome measure. *J Bone Joint Surg Br.* 2001;83B:792-4.
  126. Stiegelmar R, McKee MD, Waddell JP, Schemitsch EH. Outcome of foot injuries in multiply injured patients. *Orthop Clin North Am.* 2001;32:193-204.
  127. Dawson J, Carr A. Outcomes evaluation in orthopaedics. *J Bone Joint Surg Br.* 2001;83B:313-5.
  128. Turchin DCJ, Schemitsch EH, McKee MD, Waddell JP. Do foot injuries significantly affect the functional outcome of multiply injured patients? *J Orthop Trauma.* 1999;13:1-4.
  129. Egol KA, Dolan R, Koval KJ. Functional outcome of surgery for fractures of the ankle. *J Bone Joint Surg Br.* 2000;82B:246-9.
  130. Coughlin MJ, Saltzman CL, Nunley JA. Angular measurements in the evaluation of hallux valgus deformities: a report of the Ad Hoc committee of the American Orthopaedic Foot and Ankle Society on angular measurements. *Foot Ankle Int.* 2002;23:68-74.
  131. Coughlin MJ. The reliability of angular measurement in hallux valgus deformities. *Foot Ankle Int.* 2001;22:369-79.
  132. Covell DG, Uman GC, Manning PR. Information needs in office practice: are they being met? *Ann Intern Med.* 1985;103:596-9.
  133. Pell RE, Myerson MS, Schon LC. Clinical outcome after primary triple arthrodesis. *J Bone Joint Surg Am.* 2000;82A:47-57.
  134. Robinson J, Cook J, Purdam C, et al. The VISA-A questionnaire: a valid and reliable index of the clinical severity

- ty of achilles tendonopathy. *Br Sports Med.* 2001;35:335-41.
135. Pijnenberg CM, Van Dijk CN, Bossuyt PM, Marti RK. Treatment of ruptures of the lateral ankle ligaments: a meta-analysis. *J Bone Joint Surg Am.* 2000;82A:761-73.
  136. Metaxiotis D, Accles W, Pappas A, Doederlein L. Dynamic pedobarography (DPB) in the management of cavovarus foot deformity. *Foot Ankle Int.* 2000;21:935-47.
  137. Rowan KJ. Development and validation of a multi-dimensional measure of chronic foot pain: the Rowan Foot Pain Assessment Questionnaire (ROFPAQ). *Foot Ankle Int.* 2001;22:795-809.
  138. Acosta R, Ushiba J, Cracchiolo A. The results of primary and staged pantalar arthrodesis and tibiotalar calcaneal arthrodesis in adult patients. *Foot Ankle Int.* 2000;21:182-94.
  139. Toolan BC, Wright Quinones VJ, Cunningham BJ, Brage ME. An evaluation of the use of retrospectively acquired preoperative AOFAS clinical rating scores to assess surgical outcome after elective foot and ankle surgery. *Foot Ankle Int.* 2001;22:775-87.
  140. Bergner M. Health status measures: an overview and guide for selection. *Ann Rev Public Health.* 1987;8:191-210.
  141. Bergner M. Quality of life, health status, and clinical research. *Med Care.* 1989;27 3 Supl:S148-S56.
  142. Nunnally JC, Bernstein IH. *Psychometric theory.* 3rd ed. Toronto: McGraw-Hill; 1994.
  143. Patrick DL, Bergner M. Measurement of health status in the 1990s. *Ann Rev Public Health.* 1990;11:165-83.
  144. Ware JE Jr. The status of health assessment 1994. *Ann Rev Public Health.* 1995;16:327-54.
  145. Ware JE, Sherbourne CD. The MOS 36 item short form health survey (SF36). Conceptual framework and item selection. *Med Care.* 1992;30:473-83.
  146. Ware J, Kosinski M, Keller SD. A 12 item short-form health survey: construction of scales and preliminary tests of reliability and validity. *Med Care.* 1996;34:220-3.
  147. Guyatt GH, Naylor CD, Juniper E, et al. Users' guides to the medical literature: XII. How to use articles about health-related quality of life. *JAMA.* 1997; 277:1232-7.
  148. Spector PE. Summated rating scale construction: An introduction. Newbury Park: Sage Publications, 1992.
  149. Streiner DL, Norman GR. *Health measurement scales. A practical guide to their development and use.* Oxford/New York/Tokyo: Oxford University Press, 1994.
  150. Testa MA, Nackley JF. Methods for quality-of-life studies. *Ann Rev Public Health.* 1994;15:535-59.

**Conflicto de intereses.** El autor no ha recibido ayuda económica para la realización de este trabajo. Tampoco ha firmado ningún acuerdo por el que vaya a recibir beneficios u honorarios por parte de alguna entidad comercial por la realización de este trabajo. Por otra parte, ninguna entidad comercial ha pagado ni pagará a fundaciones, instituciones educativas u otras organizaciones sin ánimo de lucro a las que esté afiliado.