

# Evaluación de los tratamientos y técnicas de rehabilitación

**L. Martinez**  
**M. Revel**

*En 1978, el profesor Grossiord [17] justificó el desarrollo de la kinesiterapia y de la reeducación funcional como respuesta a una nueva necesidad, relacionada con el incremento de las patologías invalidantes. Llegó a la conclusión de que era necesario crear una nueva disciplina, centrada en los problemas de la reeducación funcional y de la readaptación. Desde entonces, la importancia de esta especialidad no ha hecho sino crecer y se ha impuesto en numerosos campos.*

*Sin embargo, al igual que cualquier actividad terapéutica, la rehabilitación debe someter sus tratamientos a una cuidadosa evaluación; de lo contrario, corre el riesgo de quedar relegada al rango de medicina paralela. Tanto la kinesiterapia como las terapéuticas manuales, la fisioterapia o la hidroterapia han de aportar pruebas de su eficacia mediante ensayos terapéuticos dotados de todas las garantías de una metodología rigurosa. Ésta se encuentra actualmente bien establecida, principalmente para los medicamentos, y puede ser aplicada a los ensayos terapéuticos que evalúan la rehabilitación.*

## Principios comunes a los ensayos terapéuticos

Todo ensayo terapéutico implica el empleo de tres grandes principios [15, 39]: el principio de comparación, el principio de causalidad y el principio de significación.

### Principio de comparación

El resultado de un tratamiento no se mide en términos absolutos sino, por el contrario, en relación a la ausencia de tratamiento o en comparación con otro. En razón de la variabilidad de los individuos, que es una característica de los seres vivos [36], los pacientes son distintos entre ellos y cada enfermo difiere de sí mismo a lo largo del tiempo. Con el fin de compararlos, es preciso, por tanto, seleccionar pacientes que se asemejen al menos en algunos aspectos.

### Selección de individuos [5, 39]

Se lleva a cabo mediante criterios de inclusión y exclusión estrictos y sin equívoco, lo que ha llevado a decir que los pacientes que participan en los ensayos no tienen mucho que ver con los que recibirán posteriormente el tratamiento. En realidad, tal contradicción tiene que ver con la búsqueda de la homogeneidad (lucha contra la variabilidad), lo que permite disminuir el número de pacientes que se incluirán, y con la precisión de la afección tratada. Los individuos así seleccionados habrán de parecerse al menos en

dos puntos: deberán presentar un pronóstico de evolución espontánea más o menos comparable y una sensibilidad a los tratamientos cotejados, aproximadamente similar.

Es por ello que es menester definir con exactitud el cuadro nosológico (la enfermedad y su forma clínica), la patología asociada y las características fisiológicas o demográficas de la población estudiada.

### Ensayo controlado [39]

Se habla de ensayo controlado cuando los individuos se encuentran distribuidos al menos en dos grupos. Es la única forma de obtener una información que, aunque relativa, pueda generalizarse a todos los pacientes. Un grupo recibe la terapia por validar; el otro grupo (testigo) no recibe tratamiento o es tratado con placebo o con un tratamiento de referencia.

### Principio de causalidad

Permite establecer una relación causa-efecto entre la diferencia observada entre los grupos al finalizar el ensayo y la introducción de una variable independiente de los pacientes, a saber el tratamiento. Presupone la existencia de grupos idénticos desde el inicio (de ahí la necesidad de seleccionar a los pacientes) y el respeto a otros dos requisitos metodológicos: el estudio randomizado y la prueba a ciegas.

### Estudio randomizado [21, 39]

Consiste en repartir aleatoriamente el tratamiento o los pacientes en cada uno de los grupos por tratar. Es el único procedimiento con el que se puede estar absolutamente seguro de que no existe relación alguna con las características del sujeto, la gravedad de la enfermedad ni, de forma indirecta, con el resultado del tratamiento. Impide conocer

Luc MARTINEZ: Médecin généraliste, DIU Méthodologie des essais thérapeutiques en ville, maison médicale Emile-Roux, 50, avenue Jean-Jaurès, 78390 Bois d'Arcy.

Michel REVEL: Professeur, adjoint-responsable rééducation, clinique de rhumatologie, hôpital Cochin, 27, rue du Faubourg Saint-Jacques, 75014 Paris.

el tratamiento que será administrado y previene así dos tipos de comportamiento perjudiciales para el rigor del ensayo: el de no hacer participar en el ensayo a pacientes que recibirán el tratamiento en el que no se tiene demasiada esperanza y el de aceptar a estas mismas personas en el ensayo, si se sabe que van a recibir el tratamiento que inspira más confianza. Finalmente, constituye la mejor forma de asegurarse de que el reparto de los factores no controlados (todos aquellos susceptibles de influir sobre el curso de la enfermedad y que no han sido tenidos en cuenta por los criterios de inclusión o exclusión) se lleva a cabo de manera equivalente entre los grupos estudiados.

### **Prueba a ciegas** <sup>[39]</sup>

Este término significa que la naturaleza del tratamiento es desconocida. Los pacientes, los experimentadores o el personal auxiliar pueden ser «ciegos». Los métodos así llamados se justifican por el hecho de que tanto el paciente como el médico nunca son completamente objetivos. Un experimentador, incluso si participa desde el principio, puede modificar su juicio tras el conocimiento de los primeros resultados y favorecer inconscientemente el tratamiento que prefiere. Lo mismo ocurre con el paciente.

#### *Ciego simple*

El caso en que únicamente el paciente es «ciego» carece de todo interés ya que en este procedimiento los tratamientos deben ser idénticos. Exige todas las condiciones del doble ciego sin obtener sus ventajas. Resulta inaplicable en rehabilitación por razones evidentes. En realidad, es el médico, o más exactamente el evaluador, quien debe ser «ciego». Por lo tanto, es preciso un segundo observador a la hora de incluir a los pacientes y explicar el tratamiento. Este sistema se impone si los tratamientos comparados no pueden ser enmascarados, caso más frecuente en rehabilitación.

#### *Doble ciego*

El paciente y todo el equipo auxiliar ignoran los tratamientos aplicados hasta el fin del ensayo. Este modelo experimental no suele ser aplicable en rehabilitación dado que resulta imposible encubrir los tratamientos que se comparan.

El interés de los métodos a ciegas radica en poder mantener la comparabilidad inicial de los grupos y, de esta forma, aumentar la intensidad de la relación causa-efecto al limitar la intervención de la distorsión (distorsión de selección, distorsión de evaluación, limitación de abandonos de tratamiento o mejora en el cumplimiento del tratamiento).

### **Principio de significación** <sup>[15, 35, 36, 39]</sup>

Mide la fuerza de la conclusión del ensayo. Se basa en el razonamiento estadístico y se expresa en términos de probabilidad. Tal incertidumbre está asociada a la variabilidad inherente al mundo de los seres vivos, que tiene por consecuencia las fluctuaciones del muestreo. Todo el proceso de razonamiento estadístico consiste en cuantificar el riesgo de error, cuando se llega a la conclusión de que la eficacia del tratamiento es la que justifica la diferencia encontrada entre los grupos al final del ensayo. Para llegar a ello, hay que recurrir a la hipótesis nula.

### **Hipótesis nula**

Postula la ausencia de diferencia entre los dos grupos estudiados. Si los resultados del ensayo son poco compatibles con esta hipótesis, ésta es rechazada; probablemente la diferencia no está asociada a las fluctuaciones del muestreo; el tratamiento se pone en tela de juicio, bajo reserva de que todas las exigencias ligadas al principio de comparación estén garantizadas.

### **La « $p < 0,05$ »**

Afirmar que una diferencia es significativa (es decir no debida al azar) implica necesariamente la aceptación de un riesgo de equivocarse. Este error resulta lamentable pero es inevitable. Generalmente el riesgo de error aceptado es del 5 %, lo que se expresa mediante la conocida « $p < 0,05$ ». Se habla de error de primera clase o riesgo  $\alpha$ .

### **¿Cómo llegar a la conclusión?**

— *La diferencia es estadísticamente significativa*: esta conclusión se logra a través de tablas estadísticas y significa que razonablemente la diferencia encontrada entre los grupos al final de ensayo tiene escasas probabilidades de ser el resultado del azar. No obstante, el razonamiento estadístico no lo es todo. Resta dotar a esta diferencia de un interés práctico y es el juicio clínico que puede aportarlo. Si el azar explicase la diferencia, se cometería un error  $\alpha$ , llamado de primera clase, al llegar a la conclusión de que el tratamiento es eficaz. Sin embargo, se acepta tal riesgo siempre que sea inferior al 5 %.

— *La diferencia no es significativa*: puede resultar enmascarada, por lo que se comete un error  $\beta$ , llamado de segunda clase, al no concluir en que el tratamiento es eficaz. Grandes fluctuaciones en el muestreo, número de personas demasiado escasos o un criterio de juicio excesivamente poco sensible pueden ser el origen de este fallo.

En todo caso, nunca se podrá deducir que los tratamientos son equivalentes.

## **Otros requisitos**

Se derivan en parte de lo que acaba de enunciarse.

### **Una única pregunta** <sup>[22, 39]</sup>

Un buen ensayo responde exclusivamente una sola pregunta principal, cuidadosamente elegida, claramente definida y formulada de antemano, a la cual se referirá el análisis estadístico. Multiplicar el número de interrogantes conlleva aumentar el riesgo de error de primera clase (siempre se acaba por encontrar una pregunta cuya respuesta sea significativa); ahora bien, el riesgo  $\alpha$  sólo se corre para una única pregunta.

Esto no impide proponer interrogantes secundarios en un número limitado. Pueden constituir bases de reflexión con miras a búsquedas posteriores pero no deben parasitar la respuesta a la pregunta planteada.

### **Un único criterio de juicio** <sup>[8, 39]</sup>

El razonamiento estadístico impone el recurso a un único criterio de juicio. Esto no siempre resulta sencillo en rehabilitación, donde la evaluación del estado del paciente exige tener en cuenta numerosos parámetros. Si el clínico se ve obligado a emplear varios criterios de juicio, deberá privilegiar uno de ellos, que será el principal, al cual se referirá el análisis estadístico y del cual dependerá la cantidad de pacientes que se incluirán; el resto, en cantidad limitada, constituirá los criterios secundarios.

### **Número de pacientes que se incluirán** <sup>[7, 39]</sup>

El número de pacientes que serán incluidos está determinado antes de iniciar el ensayo. Define el tamaño de la muestra y depende de cuatro parámetros:

— *sensibilidad deseada* para la comparación; mide la diferencia entre el promedio de los efectos de cada tratamiento; cuanto menor es esta diferencia, mayor cantidad de pacientes hacen falta para el estudio (a fin de combatir las fluctuaciones del muestreo, cuyo impacto es tanto mayor cuanto menor es la diferencia deseada entre los tratamientos);

— *variabilidad de los resultados*; cuanto mayor sea, más pacientes son precisos (la variancia de los resultados mide directamente la influencia de las fluctuaciones de muestreo sobre estos resultados);

— *riesgo  $\alpha$  de primera clase*; cuanto mayor sea el afán por no equivocarse, menor será  $\alpha$  pero mayor número de pacientes serán necesarios;

— *riesgo  $\beta$ , o error de segunda clase*, de no llegar a la conclusión de su eficacia cuando realmente existe; cuanto menor es, más pacientes se necesitan.

Por todos estos motivos el cálculo del tamaño de la muestra antes de empezar el ensayo posee una importancia capital. Se evita una situación adversa en la cual la probabilidad de llegar a una conclusión de eficacia (cuando el tratamiento es verdaderamente activo) apenas supera el 10 o el 20 %.

#### **Pacientes perdidos de vista** [6, 39]

Son todos aquellos pacientes para los cuales la evaluación del criterio de juicio no ha podido llevarse a cabo. A menudo constituyen un motivo de fracaso de los ensayos terapéuticos. Así pues, es menester evaluarlos de forma correcta. Cuando el número de pacientes perdidos de vista es importante y el análisis estadístico se realiza sobre los «restantes», la disminución del tamaño de la muestra hará disminuir el valor del análisis estadístico y puede conducir a una situación desfavorable de ausencia de conclusión mientras que el tratamiento había resultado eficaz. En caso de un análisis estadístico de un grupo en el que se incluye a todos los sujetos y, por tanto, también a los perdidos de vista, la actitud más segura consiste en hacer la hipótesis de la distorsión máxima, considerando que los perdidos de vista del grupo tratado son fracasos y los del grupo testigo, éxitos; en este caso, los resultados están ciertamente desvirtuados.

Por consiguiente, es preciso obstinarse en contabilizar los pacientes perdidos de vista de cada grupo y averiguar el motivo de su salida prematura del ensayo.

#### **Observancia y contaminación** [39]

Una incorrecta observancia de los tratamientos puede atenuar las diferencias de la eficacia o favorecer al menos eficaz. Este riesgo es tanto mayor cuanto más se prolongue la duración del tratamiento. Así mismo, los pacientes pueden recurrir a otras alternativas lo que va a alterar las posibilidades de llegar a concluir el ensayo.

El examen de estos principios muestra que llegar a la conclusión de la eficacia de un tratamiento conlleva un riesgo de error. Este riesgo es ineludible y está directamente asociado a la variabilidad de los individuos que componen los grupos. Puede resultar minimizado si el protocolo del ensayo sigue una metodología rigurosa.

¿Qué ocurre con la evaluación de los tratamientos y técnicas en rehabilitación? ¿Existen requisitos específicos para este tipo de ensayos? Para tratar de responder a esto han sido revisados varios trabajos con el objetivo de analizar críticamente los ensayos terapéuticos en rehabilitación.

## **Datos de la literatura médica**

Numerosos autores están de acuerdo en reconocer la falta de pruebas acerca de la eficacia de la reeducación funcional en el enfoque terapéutico de numerosas patologías osteoarticulares. Esto puede aplicarse a cualquiera de las formas de tratamiento.

### **Kinesiterapia**

En 1978, Courtyllon [10] efectuó un ensayo terapéutico con el propósito de averiguar qué técnica reeducadora se adaptaba mejor como tratamiento de la lumbalgia crónica. En 1988, Palmer [32] observó que existían pocos trabajos que evaluaran la reeducación como tratamiento de base de la parálisis espasmódica y que todos ellos presentaban errores metodológicos. En 1990, Kraag [20] constató que, sobre un total de 63 ensayos, únicamente 2 habían analizado realmente la kinesiterapia en la espondilitis anquilosante; desgraciadamente, ninguno de estos trabajos estaba randomizado.

### **Balneoterapia e hidroterapia**

En 1989, Montastruc [29] recordó la necesidad del rigor metodológico y discutió los requisitos propios de los ensayos terapéuticos que evaluaran la hidroterapia. En 1990, al constatar el escaso interés de los americanos por la hidroterapia en las patologías reumáticas, Sukenik [40] realizó un ensayo terapéutico para estimar la eficacia de los baños en agua sulfurosa y las aplicaciones de barro en la poliartritis crónica. En 1991, Collin [9] señaló que los ensayos terapéuticos que analizaban la curación termal en la lumbalgia crónica no son controlados.

### **Electroterapia**

Binder [1] y Downing [14] constataron en 1985 y 1986 respectivamente, la falta de rigor de los ensayos que estudiaban los ultrasonidos. Cada uno realizó un ensayo terapéutico; el primero, sobre la periartritis escapulo humeral y el segundo, acerca de la epicondilitis.

En 1990, Deyo [12] subrayó la ausencia de ensayos a doble ciego que valorasen la neuroestimulación transcutánea (TENS) y propuso los medios para remediarlo. Según Roques [33], el TENS es el único procedimiento electrofisioterapéutico que ha sido objeto de trabajos de validación científica recientes y a doble ciego.

### **Manipulaciones y terapéuticas manuales**

Recientemente, Di Fabio [13] ha procedido a un análisis crítico de 146 ensayos terapéuticos que evaluaban un tratamiento manual. En realidad, únicamente ha recopilado 41 trabajos que estudiaban esta forma de tratamiento de entrada. Sólo 14 de ellos poseían las características metodológicas siguientes: ensayo randomizado con simple o doble ciego, criterios de inclusión y exclusión rigurosos, descripción precisa de las modalidades terapéuticas, presentación de resultados en términos de significación estadística y cálculo del valor y del tamaño de la muestra si el ensayo no resultaba significativo. Todos analizaban técnicas manipulativas raquídeas.

### **Análisis críticos de la metodología de los ensayos terapéuticos en reeducación funcional**

En 1991, uno de los autores [26] llevó a cabo una revisión de la literatura y procedió al análisis crítico de la metodología de 35 artículos originales que evaluaban diferentes técnicas utilizadas en reeducación funcional. Sólo los estudios prospectivos (29 ensayos) fueron sometidos a un análisis crítico de la metodología empleada mediante la búsqueda de la existencia de 20 criterios metodológicos con los que se estableció una puntuación de conformidad. Los resultados mostraron una marcada falta de rigor, obteniéndose un promedio de conformidad metodológica de solamente un 32 % para los ensayos extraídos de la literatura francesa.

Junto a los principios básicos necesarios en toda evaluación terapéutica, tan poco considerados en los ensayos que evalúan la rehabilitación (ensayo controlado, randomizado, fundado en una única cuestión, con una conclusión clara, que estudia una población bien seleccionada mediante criterios de inclusión y exclusión claros y sin equívoco, con un análisis estadístico cuyas características están establecidas desde el comienzo del ensayo, teniendo en cuenta los pacientes perdidos de vista y las contaminaciones), se estimó que algunos aspectos garantes de una correcta metodología presentaban particularidades específicas que serán desarrolladas.

## Aspectos específicos de las técnicas de rehabilitación

### *Estudios retrospectivos* [30]

Plantean el delicado problema de su legitimidad, hasta el punto que algunos autores los rechazan categóricamente. Así, para Feinstein, y según Norton, «el ensayo clínico randomizado es una verdadera religión que lleva a descartar, por anticientífico, todo ensayo no randomizado».

El estudio retrospectivo procede durante su desarrollo, de forma contraria al ensayo clínico randomizado; todo el proceso se especifica con posterioridad al suceso (en este caso, el tratamiento); el sentido de la deducción se dirige entonces «desde el efecto hacia la causa». El peligro de tales estudios consiste en introducir distorsiones; es preciso reservarlos para los casos en que, por razones de orden ético o de otro tipo, el ensayo clínico randomizado no sea posible.

### *A ciegas*

Resulta difícil de poner en práctica en los ensayos terapéuticos de rehabilitación dada la imposibilidad de que los tratamientos que se comparan sean idénticos. Por lo tanto, habrá que contentarse con la prueba a ciego simple evitando que el evaluador conozca el tratamiento juzgado. Sin embargo, existe un alto riesgo de divulgación. Deyo [12] propone algunas sencillas reglas para minimizar esta distorsión: nunca realizar un ensayo cruzado a fin de evitar que el paciente saque sus propias deducciones al comparar los tratamientos; excluir del protocolo a las personas que ya hayan recibido el tratamiento que será evaluado; hacer todo lo posible para que las modalidades de desarrollo del ensayo (periodicidad de los controles, criterios de juicio) sean idénticos en ambos grupos; el terapeuta y la persona que recoja los resultados serán diferentes; en la misma línea, es necesario emplear un evaluador distinto para cada examen, lo que plantea el problema de la reproducibilidad de los exámenes y exige comparar la variabilidad entre los evaluadores.

Para asegurar el respeto al «ciego», tanto por parte del paciente como del observador, Deyo [12] pregunta al paciente si el tratamiento que acaba de recibir es el verdadero y al médico, si el paciente que ha examinado ha recibido placebo o no. A medida que el porcentaje de respuestas correctas se acerca al 100 %, menos se ha respetado la consigna de prueba «a ciegas». Si el protocolo es abierto, el 100 % de los pacientes y evaluadores deben contestar adecuadamente; si el «ciego» es perfecto, tienen el 50 % de probabilidad de hacerlo. La comparación de estos resultados con el cumplimiento y los pacientes perdidos de vista (los pacientes que reciben placebo y lo saben tienden a interrumpir el tratamiento con mayor facilidad y a consumir medicamentos reconocidos como eficaces) permite hacerse una idea real sobre el respeto de la prueba a ciegas. Sólo estas condiciones permiten que esta prueba sea verdaderamente eficaz y que la comparación conserve toda su significación.

### *Valor y tamaño de la muestra* [31]

Tres preguntas resultan fundamentales para establecer la validez de las conclusiones estadísticas de un ensayo. ¿Es suficiente la sensibilidad del estudio para deducir formalmente una relación entre dos variables? En caso afirmativo, ¿se dispone de una prueba que permita concluir en una relación de causa-efecto? ¿Cuál es la fuerza de esta relación?

Ottenbacher revisó 100 ensayos terapéuticos sobre rehabilitación funcional publicados entre 1977 y 1986; para cada uno de ellos restableció a posteriori el valor del análisis estadístico a partir del tamaño de la muestra y de una diferencia entre el promedio de los efectos que él mismo determinó según los criterios de Cohen. De esta manera demostró que sólo el 17,5 % de las 613 pruebas estadísticas empleadas en estos 100 ensayos poseían una validez superior al 80 %, lo que indica que en 8 de cada 10 casos las probabilidades de llegar a la conclusión de la eficacia del tratamiento eran muy bajas. Resulta extremadamente perjudicial, en cuanto a pérdida de tiempo y coste, que la puesta en práctica de esta sencilla norma (facilitada a través de tablas) sea tan poco tenida en cuenta [26].

### *¿Qué criterios de juicio considerar?*

Hacerse cargo de un paciente desde el punto de vista médico siempre ha estado y estará cimentado, si se quieren evitar riesgos, en la historia clínica y en los datos ofrecidos por la exploración física [42]. El análisis de los síntomas referidos por el paciente junto a la búsqueda de signos físicos asegura el diagnóstico [41]. Estos síntomas y signos se reducen a fenómenos dolorosos, rigidez o laxitud articular o a un déficit de fuerza muscular. Van seguidos de una repercusión funcional que altera la actividad diaria en mayor o menor grado. Estos elementos son los que hay que medir, si bien aparece la dificultad de convertir los fenómenos observados en datos concretos [16].

### *¿Criterio objetivo o subjetivo?*

Son preferibles los criterios físicos [26], ligados al concepto de objetividad, ya que se prestan a la cuantificación [28] y el empleo de instrumentos para medirlos parece hacerlos más fiables [34]. Esto no siempre puede ser verificado. Varios autores [11,24,42] han señalado que signos tales como el dolor, tras movimientos rotatorios del raquis, o la sensibilidad cutánea de las regiones vertebrales, paravertebrales y glúteas ofrecen un bajo coeficiente de reproducibilidad y no constituyen buenos criterios de juicio.

Rothstein [34] ha recalado estas confusiones y ha definido la medida objetiva en términos de variancia: una medición es objetiva si las variancias inter e intraobservadores resultan mínimas. En otras palabras, esto significa que estas mediciones son independientes del observador. Pero, contrariamente a lo afirmado por Bohannon [29], en la opinión de los autores, el examinador puede ser el instrumento que realice la medición si su fiabilidad está demostrada. En esta hipótesis se hace hincapié sobre la propia medición y la objetividad califica la calidad de la misma. Una medida subjetiva se define entonces como aquella fuertemente asociada al observador, con el consiguiente riesgo de introducir una distorsión sistemática y no aleatoria, que altere la calidad de la medición [19].

Sin embargo, objetivo y subjetivo pueden referirse a la naturaleza del fenómeno a medir. En este caso, una medida subjetiva evalúa un aspecto de la subjetividad del paciente. Se trata de una sensación que sólo puede ser experimentada por él. Es lo que se hace habitualmente cuando se cuantifica el dolor, la fatiga, el grado de satisfacción del paciente o la repercusión sobre las actividades diarias. Aquí, la medida subjetiva se caracteriza por el hecho de depender en gran medida de la percepción por parte del enfermo.

Ya se trate de criterios que midan parámetros objetivos o que evalúen un aspecto de la subjetividad de los pacientes, debe tenderse hacia el empleo de mediciones objetivas, es decir, escasamente asociadas a las percepciones del evaluador. Es menester, además, asegurarse de su capacidad para evaluar mediante pruebas de correlación un estado patológico dado. En este caso debe realizarse una gran tarea de búsqueda.

#### *Escalas de evaluación*

Constituyen sistemas sencillos y poco costosos que permiten caracterizar fenómenos complejos, mayormente relacionados con las percepciones del paciente (escalas de comportamiento, que valoran una discapacidad funcional, o escalas de percepción, que miden la satisfacción del paciente).

#### • Diferentes tipos de escalas

Dependiendo del número de dimensiones que tienen en cuenta, se distinguen:

— *escalas unidimensionales*, que aprecian globalmente en intensidad la dimensión estudiada; poseen el inconveniente de reducir a una sola dimensión los fenómenos complejos como el dolor crónico o la discapacidad funcional;

— *escalas multidimensionales*, que distinguen varias dimensiones del fenómeno estudiado, dimensiones que serán medidas con criterios o indicadores específicos. Se citarán el «Mac Gill pain questionnaire» y el «sickness impact profile».

Según cuál sea el objetivo de lo que se mide, se dividen en: *escalas de estado*, concebidas para cuantificar el estado del paciente en un momento dado [25], y *escalas de transición*, que valoran directamente el cambio de estado del paciente.

Según Launois [25], las primeras están destinadas a permitir la segmentación de la población, y su puesta en práctica en el marco de un ensayo randomizado está condenada de antemano al fracaso puesto que no han sido ideadas para registrar el impacto de un tratamiento.

Según el sistema de respuesta de los pacientes, se distinguen [4]:

— *la escala verbal simple (EVS)* en 4 ó 5 categorías ordenadas, que es una de las escalas más corrientemente utilizadas; se asigna una puntuación a cada categoría; su principal ventaja es la simplicidad y su inconveniente es la falta de sensibilidad en razón del número limitado de categorías; es posible combinar los adjetivos que describen las diferentes categorías con adverbios de intensidad para aumentarlas;

— *la escala numérica*, en la que el paciente responde oralmente o por escrito asignando una puntuación entre 0 y 100; el grado de alivio puede ser expresado en porcentaje;

— *la escala análoga (o analógica) visual (EAV)*, inicialmente empleada para evaluar estados subjetivos (bienestar, depresión), fue aplicada por Huskisson para el dolor y posteriormente en otros ámbitos, como es el caso de las escalas de evaluación de la lumbalgia de Oswestry (que valoran la intensidad del dolor y la capacidad de levantar pesos, andar y desplazarse de los pacientes con dolor lumbar), donde cada dimensión es cuantificada sobre una EAV; la escala se presenta en forma de una línea horizontal de 100 mm cuyos límites están definidos por las intensidades extremas de la magnitud estudiada; el paciente dibuja un trazo sobre la línea según estime su afectación; la distancia, medida al milímetro, sirve de índice numérico; el interés de la EAV en relación a las dos precedentes se basa en su mayor sensibilidad, relacionada a una mayor posibilidad de elección sobre la línea continua; sin embargo, su inconveniente radica en la dificultad de comprensión para determinadas personas.

#### • Construcción de una escala [23, 28]

Todo estudio de la mejora de los pacientes en reeducación funcional debe definir previamente el universo objeto del análisis y las dimensiones de este universo que serán cuantificadas. De una manera global, puede conocerse la calidad de vida por medio de dimensiones como el buen estado físico, la repercusión sobre las actividades diarias, la capacidad de trabajo, el estado psíquico o el dolor. La elección de las dimensiones puede hacerse mediante estudio previo de la calidad de vida sobre una muestra de pacientes [3] o por consenso entre expertos. Las dimensiones son medidas mediante criterios o indicadores que varían dependiendo del objetivo fijado: biomédicos (limitados a los trastornos clínicos), del comportamiento (para evaluar la discapacidad funcional) o de la percepción (basados esencialmente en las satisfacciones o insatisfacciones de los individuos).

A continuación han de marcarse las medidas. Si bien esto es relativamente fácil cuando se refiere a indicadores físicos (talla, medición angular de una articulación, medición de la fuerza mediante un dinamómetro, medición de la presión intradiscal mediante manómetros), el recurso a indicadores de percepción plantea muchas más dificultades. En este caso, es preciso anteponer la reproductibilidad de la medición [16, 23] especificando correctamente aquello que se desea observar, el método empleado en la observación mediante una descripción detallada del instrumento utilizado y sus condiciones de uso y explicando su forma de funcionamiento.

#### • Puntuación global

La utilización de una única puntuación por adición de la nota de cada uno de los diferentes síntomas, con o sin ponderación, aporta la ventaja teórica de reforzar la potencia del análisis estadístico ya que sólo recurre a un único criterio de juicio [26, 39]. Puede reprocharse a estos índices su arbitrariedad y su oscuro significado clínico. Para Merbitz [27], la totalización de las puntuaciones de cada apartado de una escala carece de significado coherente; todo lo que se sabe es que la puntuación final es mayor que cada una de las puntuaciones individuales. Por otra parte, una misma puntuación global para dos pacientes distintos no tiene necesariamente el mismo significado (puntuaciones altas y bajas en diferentes apartados); este escollo puede salvarse mediante numerosos métodos de análisis factorial [38] o gracias a sistemas de estandarización del tipo Gosset-Student.

Sin embargo, la experiencia demuestra que las puntuaciones globales han sido utilizadas con frecuencia para juzgar la eficacia de los tratamientos y Hamilton [18] destaca que, desde su publicación hace 24 años, el índice de Barthel no ha fracasado.

#### *Validación de los criterios de juicio*

#### • Cualidades requeridas [18, 37, 39]

Habitualmente, un buen criterio de juicio debe poseer las siguientes cualidades. La *pertinencia*, o «content value» según los anglosajones. Este concepto implica que todas las dimensiones contenidas en el concepto medido son tenidas en cuenta por el instrumento de medición. A título de ejemplo, un test sobre el buen estado físico que sólo considere la resistencia presentaría una escasa validez de contenido porque omitiría la medición de otras dimensiones importantes como la fuerza, la velocidad, la destreza o la motivación. La selección de los enunciados del examen se obtiene, ya sea mediante consulta a expertos, o por selección de las declaraciones de los propios pacientes, lo que mejora la exhaustividad. La *receptividad* evalúa la formulación de un cuestionario, condiciona el rigor de las respuestas. También se denomina validez de superficie o «face

value». La *fiabilidad* o *reproductibilidad* tiene por objeto eliminar las redundancias y garantizar una correcta reproductibilidad de las medidas inter e intraobservadores. La *sensibilidad* es la capacidad del instrumento para detectar cambios de pequeña amplitud en los pacientes. La *validez* propiamente dicha se asegura de que el instrumento mida efectivamente aquello para lo que ha sido creado. En ausencia de patrón, esta validez sólo puede ser obtenida por comparación (mediante tests de correlación) con otros indicadores que exploren el mismo campo.

#### • Postura de Feinstein [16]

Feinstein considera que el retraso de los datos clinimétricos está relacionado a los hábitos bioestadísticos. La ausencia de patrón ha conducido a una proliferación de los criterios de validez (validez de construcción, de convergencia y de divergencia), de los que sólo uno merece la atención de los autores: la «validez aparente» o «sensibilidad clínica», que apela a la sensatez y a un profundo conocimiento clínico de la enfermedad. Es preciso asociarle la reproductibilidad, requisito indispensable de un buen instrumento.

Los índices estadísticos no son suficientes para establecer un buen criterio de juicio. Éste debe poseer además una adecuada sensibilidad clínica. Para ello ha de estar adaptado a su finalidad clínica y a la situación considerada; debe ser de fácil utilización y sus resultados han de ser comprensibles; las instrucciones de uso deben ser precisas; debe permitir las discriminaciones deseadas; es necesario que existan componentes aceptables, con peso relativo, basados en la calidad científica de los datos netos.

\*  
\*\*

*La evaluación de los tratamientos y técnicas en rehabilitación apenas ha comenzado, y la demanda de pruebas de eficacia en esta rama de la medicina se hace cada vez más acuciante. No obstante, las reglas metodológicas para la evaluación terapéutica existen y son enormemente utilizadas en los medicamentos. El retraso existente en rehabilitación obedece a varias razones. El desconocimiento de la necesidad del razonamiento estadístico para evaluar la eficacia de un tratamiento constituye, sin lugar a dudas, uno de los motivos. A ello hay que añadir la dificultad para juzgar el estado clínico del paciente; ya no es posible contentarse exclusivamente con parámetros físicos: es preciso considerar la repercusión funcional de la enfermedad y el grado de satisfacción de los pacientes. La falta de herramientas fiables y válidas para medir estos dos parámetros hace la tarea más complicada y constituye un reto para el futuro. Para conseguirlo [16], han de analizarse las respuestas a cuatro preguntas. Dos de ellas se plantean al paciente antes del tratamiento: «¿cómo está usted?» y «¿qué puedo hacer por usted?». Las otras dos tienen lugar con posterioridad: «¿cómo está usted actualmente?» y «¿todo ha salido bien?». Entonces los tratamientos y las técnicas empleadas en rehabilitación podrán adquirir una respetabilidad científica.*

Cualquier referencia a este artículo debe incluir la mención: MARTINEZ L. et REVEL M. – Evaluation des traitements et des techniques en rééducation. – Encycl. Méd. Chir. (Elsevier, Paris-France), Kinésithérapie-Rééducation fonctionnelle, 26-005-B-10, 1994, 6 p.

## Bibliografía

- [1] BINDER A, HODGE G, GREENWOOD AM, HAZLEMAN BL, PAGE THOMAS DP. Is therapeutic ultrasound effecting in treating soft tissue lesions? *Br Med J* 1985; 290: 512-514
- [2] BOHANNON RW. Objective measures. *Phys Ther* 1989; 69: 590-593
- [3] BOURREAU F, LUU M, DOUBRERE JF, GAY C. Elaboration d'un questionnaire d'autoévaluation de la douleur par liste de qualificatifs. Comparaison avec le McGill Pain Questionnaire de Melzack. *Thérapie* 1984; 39: 119-129
- [4] BOURREAU F, LUU M, GAY C, DOUBRERE JF. Les échelles d'évaluation de la symptomatologie douloureuse chronique. *Méd Hyg* 1982; 40: 3797-3805
- [5] BOUVENOT G. Quels patients pour nos essais? In: Inserm Ed. Le quotidien du médecin. Essais thérapeutiques, mode d'emploi. INSERM. Paris. 1990; pp 33-36
- [6] BOUVENOT G. Les perdus de vue. In: Inserm Ed. Le quotidien du médecin. Essais thérapeutiques, mode d'emploi. INSERM. Paris. 1990; pp 131-134
- [7] BOUVENOT G. Combien de sujets? In: Inserm Ed. Le quotidien du médecin. Essais thérapeutiques, mode d'emploi. INSERM. Paris. 1990; pp 65-67
- [8] BOUVENOT G, DOYON F, ESCHWEGE E. Le critère de jugement. Un enjeu capital. In: Inserm Ed. Le quotidien du médecin. Essais thérapeutiques, mode d'emploi. INSERM. Paris. 1990; pp 47-50
- [9] COLLIN JF, CONSTANT F, GUILLEMIN F, BOULANGÉ M. Evaluation de l'efficacité de la cure thermique de Bains-Bains sur les lombalgies dégénératives chroniques. *Presse therm climat* 1991; 128: 29-33
- [10] COURTILLON A. Approche thérapeutique dans la réadaptation des lombalgies chroniques: essai comparatif de 3 techniques de rééducation lombaire. *Mém CES Rhumatologie*. Paris. 1978; 98 p
- [11] DEYO RA, RAINVILLE J, KENT DL. What can the history and physical examination tell us about low back pain? *JAMA* 1992; 268: 760-765
- [12] DEYO RA, WALSH NE, SCHOENFELD LS, RAMAMURTHY S. Can trials of physical treatments be blinded? *Am J Phys Med Rehabil* 1990; 69: 6-10
- [13] DI FABIO RP. Efficacy of manual therapy. *Phys Ther* 1992; 72: 853-864
- [14] DOWNING DS, WEINSTEIN A. Ultrasound therapy of subacromial bursitis. A double-blind trial. *Phys Ther* 1986; 66: 194-199
- [15] ESCHWEGE E. Le trépied méthodologique de l'essai thérapeutique. In: Inserm Ed. Le quotidien du médecin. Essais thérapeutiques, mode d'emploi. INSERM. Paris. 1990; pp 9-11
- [16] FEINSTEIN A. La démesure de la mesure: introduction à la clinimétrie. In: Launois R, Régnier F. eds. *Décision thérapeutique et qualité de vie*. John Libbey Eurotext. Paris. 1992; pp 103-112
- [17] GROSSIORD A, PANNIER S. Kinésithérapie et rééducation fonctionnelle. *Encycl Méd Chir* (Paris-France), Kinésithérapie-Rééducation fonctionnelle, 26005 A10 1978; 11 p
- [18] HAMILTON BB, GRANGER CV. Letter to the editor. *Arch Phys Med Rehabil* 1989; 70: 861-862
- [19] JETTE AM. Measuring subjective clinical outcomes. *Phys Ther* 1989; 69: 580-584
- [20] KRAAG G, STOKES B, GROH J, HELEWA A, GOLDSMITH C. The effects of comprehensive home physiotherapy and supervision on patients with ankylosing spondylitis. A randomized controlled trial. *J Rheumatol* 1990; 17: 228-233
- [21] LACROUX A, BOUVENOT G. Le tirage au sort est-il justifié? In: Inserm Ed. Le quotidien du médecin. Essais thérapeutiques, mode d'emploi. INSERM. Paris. 1990; pp 13-15
- [22] LACROUX A, ESCHWEGE E. Quelle est la question? In: Inserm Ed. Le quotidien du médecin. Essais thérapeutiques, mode d'emploi. INSERM. Paris. 1990; pp 29-31
- [23] LAUNOIS R. La qualité de vie: panorama et mise en perspective. In *Décision thérapeutique et qualité de vie*. Launois R, Régnier F. John Libbey Eurotext. Paris. 1992; pp 3-24
- [24] MCCOMBE PF, FAIRBANK JC, COCKERSOLE BC, PYNSENT PB. Reproducibility of physical signs in low-back pain. *Spine* 1989; 14: 908-918
- [25] MACKENZIE CR, CHARLSON ME. Standarts for the use of ordinal scales in clinical trials. *Br Med J* 1986; 292: 40-43
- [26] MARTINEZ L. Les essais thérapeutiques en médecine de rééducation: difficultés méthodologiques (Mém: DIJ Méthodologie des essais thérapeutiques en ville). Paris. 1991; 35 p
- [27] MERBITZ C, MORRIS J, GRIP JC. Ordinal scales and foundations of misinference. *Arch Phys Med Rehabil* 1989; 70: 308-312
- [28] MILLION R, HALL W, NILSEN KH, BAKER RD, JAYSON MI. Assessment of the progress of the back-pain patient. *Spine* 1982; 7: 204-212
- [29] MONTASTRUC JL, MONTASTRUC P. La pharmacologie clinique: un modèle pour l'étude scientifique du thermalisme? *Presse Therm Climat* 1989; 126: 5-7
- [30] NORTON BJ, STRUBE MJ. Some cautionary comments on the use of retrospective designs to evaluate treatment efficacy. *Phys Ther* 1988; 68: 1374-1377
- [31] OTTENBACHER KJ, BARRETT KA. Statistical conclusion validity of rehabilitation research. A quantitative analysis. *Am J Phys Med Rehabil* 1990; 69: 102-107
- [32] PALMER FB, SHAPIRO BK, WACHTEL RC et al. The effects of physical therapy on cerebral palsy. A controlled trial in infants with spastic diplegia. *N Engl J Med* 1988; 318: 803-808
- [33] ROQUES CF. Electrothérapie des lombalgies. *RMO* 1992; 27: 6-12
- [34] ROTHSTEIN JM. On defining subjective and objective measurements. *Phys Ther* 1989; 69: 577-579
- [35] SCHWARTZ D. «p». In: Inserm Ed. Le quotidien du médecin. Essais thérapeutiques, mode d'emploi. INSERM. Paris. 1990; pp 21-23
- [36] SCHWARTZ D. La statistique dans les sciences du vivant. Dossier documentaire. INSERM. Paris. 1991; 27 p
- [37] SIM J, ARNELL P. Measurement validity in physical therapy. *Phys Ther* 1993; 73: 102-110
- [38] SPRIET A, FERMANIAN J, SIMON P. L'utilisation des échelles d'évaluation en psycho-pharmacologie. *Encéphale* 1978; 4: 119-129
- [39] SPRIET A, SIMON P. Méthodologie des essais cliniques des médicaments. Laboratoires Hoescht. Puteaux. 1986; 231 p
- [40] SUKENIK S, BUSKILA D, NEUMANN L, KLEINER-BAUMGARTEN A, ZIMLICHMAN S, HOROWITZ J. Sulphur bath and mud pack treatment for rheumatoid arthritis at the Dead Sea area. *Ann Rheum Dis* 1990; 49: 99-102
- [41] TROISIERS O. Diagnostic clinique en pathologie ostéo-articulaire. Masson. Paris. 1990; 116 p
- [42] WADDELL G, MAIN CJ, MORRIS EW et al. Normality and reliability in the clinical assessment of backache. *Br Med J* 1982; 284: 1519-1523