

## Sensibilidad al cambio de las medidas de desenlace

María Jesús García de Yébenes Prous<sup>a</sup>, Francisco Rodríguez Salvanés<sup>b</sup> y Loreto Carmona Ortells<sup>a</sup>

<sup>a</sup>Unidad de Investigación. Fundación Española de Reumatología. Madrid. España.

<sup>b</sup>Agencia Laín Entralgo. Madrid. España.

En la investigación médica, y en especial en el ámbito de la reumatología, hay gran interés sobre el concepto de sensibilidad al cambio de un instrumento de medida como posible reflejo de modificaciones en la situación clínica del paciente. Sin embargo, la terminología de este concepto y su metodología de análisis están rodeadas de confusión y falta de consenso.

Se presenta un trabajo de revisión sobre el concepto y el análisis de la sensibilidad al cambio teniendo en cuenta tanto las características de la sensibilidad como el tipo de diseño y cambio previsible en la muestra en estudio.

**Palabras clave:** Sensibilidad al cambio. Medidas de desenlace. Estadísticos de sensibilidad al cambio.

### Responsiveness of Outcome Measures

In medical research, particularly in the field of rheumatology, there is great interest about the concept of responsiveness of outcome measures as a sign of changes in the patient's health status. However, the terminology surrounding this concept and the methods of analysis are confusing and lacking in consensus.

We present a review about the concept and analysis of responsiveness taking into account both, the characteristics of the responsiveness and the type of design and predictable change in the sample being studied.

**Key words:** Responsiveness. Outcome measures. Responsiveness statistics.

### Sensibilidad al cambio

En general, los clínicos y los investigadores evalúan la calidad de las escalas de medición en función de su fiabilidad y su validez. Sin embargo, la sensibilidad al cambio (*responsiveness*) también debe considerarse una característica fundamental de los instrumentos de evaluación, diseñados para medir un cambio longitudinal a lo largo del tiempo<sup>1</sup>. En la literatura no hay consenso sobre el concepto de sensibilidad al cambio de un instrumento de medida ni sobre la forma en que debe cuantificarse. En una revisión de la literatura médica, Terwee et al<sup>2</sup> encontraron distintas definiciones y 31 formas diferentes de medir la sensibilidad al cambio. Esta falta de consenso ha originado una gran proliferación de estadísticos, y no es infrecuente que algunos investigadores utilicen simultáneamente varios en el mismo artículo<sup>3,4</sup>, lo que dificulta o incluso impide la comparación de medidas entre diferentes estudios<sup>5</sup>. La incertidumbre sobre el tipo de diseño y los métodos de análisis puede deberse, en gran medida, a la ausencia de un estándar de cambio en el estado de salud<sup>6</sup>.

La sensibilidad al cambio es el grado con que se obtienen diferentes resultados en aplicaciones repetidas del mismo instrumento cuando se ha producido un cambio real en el estado de salud<sup>7</sup>. Por lo tanto, puede definirse como la capacidad de un instrumento para detectar un cambio. Esta característica despierta mucho interés, ya que el cambio en una medida puede reflejar una modificación de la situación clínica del paciente, lo que es importante para los estudios de intervención.

El estudio de esta dimensión requiere de un estándar que indique el cambio clínico. El método tradicionalmente utilizado para evaluar las características del cambio es la evaluación retrospectiva y general del cambio producido<sup>7,8</sup>. Este método consiste en que durante las visitas de seguimiento, además de volver a realizar la prueba en cuestión, se pide al paciente su impresión general del cambio producido. Este único ítem de evaluación retrospectiva general se emplea después para valorar la capacidad del instrumento para detectar un cambio<sup>9</sup>. No obstante, esta manera de proceder tiene varios inconvenientes: *a*) las propiedades métricas (coherencia interna y validez) de un único ítem de evaluación retrospectiva general son inferiores a las del instrumento o cuestionario de múltiples ítems en estudio; *b*)

Correspondencia: Dra. M.J. García de Yébenes y Prous.  
Unidad de Investigación. Fundación Española de Reumatología.  
Marqués del Duero, 5, 1.º A. 28001 Madrid. España.  
Correo electrónico: mjesus.garciadeyebenes@ser.es

Manuscrito recibido el 27-6-2008 y aceptado el 24-7-2008.

desde un punto de vista psicológico, la valoración retrospectiva general es difícil y subjetiva, y c) la utilización de la valoración retrospectiva general se basa en la presunción de que es una medida independiente del instrumento en estudio, por lo que los errores de medición no están correlacionados. Sin embargo, cualquier exacerbación o remisión en el momento de la observación puede influir en la valoración del paciente. La correlación del error de medida entre la valoración retrospectiva general y la prueba sobrestima la relación real entre ambas. Por consiguiente, la evaluación de la sensibilidad al cambio a base de una valoración retrospectiva general tiene escaso valor para estudiar la capacidad de un instrumento de detectar un cambio<sup>4,8</sup>.

Una alternativa a la valoración retrospectiva general es la valoración pronóstica del cambio, que consiste en una declaración a priori de las características del cambio que se va a producir en la muestra. Este procedimiento no está sometido a errores de recuerdo o de correlación, pero depende de la capacidad del evaluador de estimar de forma exacta la extensión del cambio que puede producirse. Se ha utilizado, por ejemplo, en ensayos aleatorizados en los que se comparan intervenciones de eficacia conocida con placebo o en estudios de cohortes en los que se emplea una variable pronóstica conocida para clasificar a los pacientes en grupos que previsiblemente cambiarán de diferente manera<sup>8</sup>.

La sensibilidad al cambio depende de la población de pacientes estudiada y del escenario en que se aplica el instrumento de medición. Por ejemplo, la variabilidad de las puntuaciones de cambio será mayor en una población heterogénea que en una homogénea; del mismo modo, cuando la intervención sea muy eficaz en algunos pacientes pero no en otros el cambio será más importante que cuando sea poco eficaz en todos los pacientes, aunque la media del cambio sea la misma en ambos casos<sup>2</sup>. Por consiguiente, la elección del método analítico y de los respectivos coeficientes de sensibilidad al cambio depende fundamentalmente de las características de la muestra, sobre todo del tipo de diseño y del cambio esperable<sup>8,10</sup>:

– Un único grupo: cambio homogéneo. La muestra está formada por un único grupo en el que los pacientes previsiblemente cambiarán más o menos de la misma forma entre dos momentos. Los coeficientes que se utilizan se basan en el cambio homogéneo entre los pacientes.

– Un único grupo: cambio heterogéneo. La muestra también está formada por un único grupo pero, a diferencia del anterior, es previsible que los pacientes cambien de diferente forma entre ellos. Los coeficientes se basan en análisis de correlación.

– Diferentes subgrupos: cambio mixto. En ese caso la muestra está constituida por dos o más subgrupos de pacientes que cambian de diferente forma entre dos momentos. Comparte características de los dos anterior-

es: el cambio será diferente entre los subgrupos de pacientes definidos en la muestra, por lo que sería un diseño de tipo 2, pero además, dentro de cada grupo, el cambio será previsiblemente homogéneo, lo que le confiere características del primer diseño. Los coeficientes se basan en diferencias de cambio entre subgrupos.

Algunos autores utilizan otros sistemas de clasificación de la sensibilidad al cambio. En este contexto, Husted et al distinguen dos grandes tipos de sensibilidad al cambio: interna y externa<sup>5,11</sup>.

La sensibilidad al cambio interna se define como la capacidad de una medida para cambiar en un lapso determinado; se refiere a la posibilidad de detectar cualquier cambio de tipo estadístico. Por ejemplo, se estudia un único grupo de pacientes a los que se evalúa antes y después de aplicar un tratamiento eficaz. La sensibilidad al cambio dependerá tanto del tratamiento empleado como de la medida de resultado utilizada para determinar la eficacia terapéutica.

La sensibilidad al cambio externa se define como el grado con que los cambios en una medida en tiempo se relacionan con cambios correspondientes en una medida estándar del estado de salud. Esta dimensión de la sensibilidad al cambio está asociada al concepto de relevancia clínica y consiste en la propiedad de una medida para capturar un cambio clínicamente importante. Al contrario que en la interna, el interés fundamental no está en la propia medida, sino en la relación entre el cambio en la medida y el cambio en el estándar externo. Si esta relación es importante, la medida captura adecuadamente los cambios en el estándar externo. Se acepta que los cambios en el estándar son una indicación de una modificación de la situación del paciente. Por consiguiente, que el cambio en la medida sea capaz de capturar el cambio en el estándar podría indicar una modificación en la situación clínica del sujeto. La sensibilidad al cambio externa dependerá únicamente de la elección del estándar externo, y no del tratamiento en estudio. Por consiguiente, este tipo de sensibilidad al cambio será una propiedad del instrumento de medida. En la tabla 1 se presenta una clasificación de la sensibilidad al cambio en función de sus características y el tipo de diseño/cambio previsible.

A continuación se presentan los estadísticos más utilizados para evaluar la sensibilidad al cambio, tanto en función del diseño del estudio y las características del cambio previsible como por la clasificación en sensibilidad al cambio interna y externa.

### Cambio homogéneo

Como ya se ha comentado, este diseño y su correspondiente análisis se basan en la premisa de que la muestra está formada por un único grupo de pacientes que pre-

**TABLA 1. Sensibilidad al cambio según sus características y el tipo de diseño**

Diseño/ tipo de cambio	Características	
	Interna	Externa
Homogéneo	Prueba t datos relacionados Tamaño del efecto estandarizado Respuesta media estandarizada (SMR) Estadístico de Guyatt	
Contraste entre subgrupos		Curvas ROC ANOVA medidas repetidas S de Norman
Contraste entre pacientes		Correlación Modelos de regresión

visiblemente cambiarán de la misma forma durante el periodo de estudio. Lo importante no es los factores de los que depende el cambio (historia natural o aplicación de un tratamiento eficaz), sino que la magnitud del cambio es homogénea entre los pacientes.

Este diseño mediría la anteriormente mencionada “sensibilidad al cambio interna”, ya que, en realidad, lo que se evalúa es la capacidad de la medida para cambiar durante un lapso concreto en una muestra de pacientes que mejoran, por ejemplo, tras aplicarles un tratamiento de eficacia conocida<sup>5</sup>.

Los estadísticos más utilizados se basan en el cambio medio o general de un grupo, y entre ellos hay que citar los siguientes.

### Prueba de la t para datos relacionados

El estadístico t prueba la hipótesis de ausencia de cambio en la respuesta media de una medida entre dos momentos. Puesto que se trata de un diseño con medidas repetidas en el mismo sujeto, se utiliza la t para datos relacionados.

$$t = \text{media de la diferencia} / (\text{DE}_{\text{diferencia}} / \sqrt{n})$$

El estadístico t se centra en la significación estadística del cambio observado, que depende, evidentemente, de la magnitud del cambio, pero también del tamaño de la muestra y de la variabilidad de la medida. Esto hace que no sea un estadístico potente para evaluar sensibilidad al cambio. Desde un punto de vista metodológico, es más adecuado utilizar coeficientes en los que intervengan tanto la magnitud del cambio como su variabilidad.

### Estadísticos basados en el tamaño del efecto

Al contrario que las pruebas t, los estadísticos basados en el tamaño del efecto suministran información directa sobre la magnitud del cambio, expresado en forma de variación. Por lo tanto, los estadísticos de tamaño del efecto miden la relación entre la magnitud del cambio (señal) y la variabilidad (ruido de fondo).

#### Tamaño del efecto estandarizado

Un estadístico de este grupo, muy utilizado, es el tamaño del efecto estandarizado o relación entre la media de las diferencias entre las puntuaciones basal y al seguimiento dividida por la desviación estándar (DE) de la medida basal<sup>3</sup>.

$$\text{Tamaño del efecto} = \text{cambio medio} / \text{DE}_{\text{medida basal}}$$

Un tamaño del efecto de 0,20 indica que el cambio es de alrededor de una quinta parte de la DE de la medida basal y se define como pequeño. Valores de 0,50 se consideran cambio moderado y a partir de 0,80, cambio importante. Una limitación del tamaño del efecto es que no podemos saber si refleja un cambio real o la variabilidad de la puntuación basal.

#### Respuesta media estandarizada (SMR)

Es otro estadístico basado en el tamaño del efecto. Se calcula dividiendo el cambio medio por su DE, lo que refleja la variabilidad de las puntuaciones de cambio. Por lo tanto, si el cambio tiene una elevada variabilidad respecto a su media, se obtendrá un valor bajo de SMR.

$$\text{SMR} = \text{cambio medio} / \text{DE}_{\text{cambio}}$$

La respuesta media estandarizada es independiente del tamaño de la muestra y, además, tiene en cuenta la variabilidad del cambio, por lo que es el estadístico más adecuado para el estudio de la sensibilidad en este tipo de diseños. Al igual que con el tamaño del efecto, los valores de 0,20, 0,50 y 0,80 indican instrumentos con escasa, moderada y elevada sensibilidad al cambio<sup>5,8</sup>.

El valor de la SMR debe acompañarse de un intervalo de confianza. Algunos autores calculan ese intervalo asumiendo que las diferencias de puntuación siguen una distribución normal de media 0 y  $\text{DE} = 1 / \sqrt{n}$ , aunque otros utilizan métodos de remuestreo como el Jackknife<sup>12</sup>.

### Estadístico de Guyatt

El estadístico de Guyatt, denominado estadístico de sensibilidad al cambio (*responsiveness statistic*), también se basa en el tamaño del efecto y consiste en el cociente entre la mínima diferencia clínicamente importante (MCID) y el error de la media cuadrática en un análisis de la variancia de medidas repetidas en pacientes clínicamente estables (MSE).

$$\text{Guyatt} = \text{MCID} / \sqrt{2 \times \text{MSE}}$$

En el caso de que sólo haya dos observaciones (antes y después de una intervención), el denominador será la DE de las puntuaciones de cambio individual en los pacientes estables.

La MCID se define como la diferencia más pequeña entre la puntuación basal y la de seguimiento que se asocia a un efecto clínico importante en un grupo de pacientes. En general, la MCID se establece relacionando los cambios en la medida con los producidos en un estándar de cambio clínico, como la propia valoración del paciente. La MCID refleja la magnitud de cambio en la medida asociada con una definición arbitraria del cambio mínimo importante en el estándar externo. Existen múltiples definiciones de MCID; una de las más utilizadas es la diferencia de cambio promedio entre pacientes que muestran alguna mejoría y los que no muestran cambios en su estado de salud. El denominador ajusta los posibles cambios espurios debidos a error de medida o efectos de aprendizaje.

El estadístico de Guyatt es una forma de detectar modificaciones en la medida de desenlace diferentes de los cambios aleatorios presentados por los pacientes que no muestran mejorías detectables<sup>6</sup>. De esta forma, una medida será sensible al cambio si es capaz de detectar MCID superiores a cualquier cambio espurio<sup>5,13</sup>. Este estadístico es poco utilizado, al menos en reumatología, debido a las dificultades inherentes a la definición del cambio mínimo clínicamente importante<sup>13</sup>.

La sensibilidad al cambio interna, o capacidad para detectar un cambio estadístico en la medida, puede verse afectada por diferentes parámetros estructurales como el tipo de escala, el sistema de puntuación y el número de ítems relacionados con la "señal" y con el "ruido". Un número elevado de ítems tiende a aumentar la sensibilidad siempre y cuando no sean redundantes. Las escalas continuas que cubren todo el espectro de un desenlace, desde formas leves a graves, generalmente evitan los efectos suelo y techo y aumentan la sensibilidad. Las escalas con sistemas de puntuación graduales también tienden a tener mayor sensibilidad al cambio que las dicotómicas. Por último, la fiabilidad es un factor que tener en cuenta, ya que interviene en la magnitud del denominador de los estadísticos basados en el tamaño del efecto<sup>11</sup>.

A excepción del de Guyatt, ninguno de estos estadísticos relaciona el cambio en la escala en cuestión con el producido en una medida de la situación clínica. Cualquier cambio observado se atribuye a una modificación del estado del paciente pero, en realidad, pueden observarse cambios estadísticamente significativos en la medida sin que se haya producido una modificación relevante en la situación clínica del sujeto<sup>4,5</sup>. Por consiguiente, sólo sirven para evaluar la capacidad intrínseca o interna del instrumento para cambiar. Por otra parte, aunque el estadístico de Guyatt se relaciona con un estándar de cambio clínico, al ser una muestra de pacientes estables no es posible diferenciar entre distintas cantidades de cambio (mejoría, empeoramiento). Estas limitaciones hacen del diseño homogéneo (sensibilidad al cambio interna) el más débil, ya que no permite discriminar entre distintas magnitudes de cambio<sup>8</sup>.

### Diseño mixto: contraste entre subgrupos

Este diseño comparte características de los otros dos. La muestra está formada por subgrupos de pacientes que cambian de diferente manera (cambio heterogéneo), pero dentro de cada subgrupo el cambio entre pacientes es uniforme (cambio homogéneo). Sería equivalente a la sensibilidad al cambio externa.

En este tipo de diseño pueden utilizarse distintos coeficientes de cambio.

### Curvas ROC

La utilización de curvas de las características operativas del receptor (ROC) como método de evaluación de la sensibilidad al cambio en reumatología fue propuesta inicialmente por Deyo et al<sup>14</sup> en 1986.

Una medida de desenlace puede describirse, de forma similar a una prueba diagnóstica, por su capacidad para identificar correctamente a los individuos que presentan un cambio clínico importante. Para poder realizar este análisis, es preciso disponer de un estándar externo de cambio. De esta forma, la sensibilidad al cambio se describe en términos de sensibilidad (probabilidad de que la medida clasifique correctamente a los pacientes que mostraron cambios en un indicador externo de cambio) y especificidad (probabilidad de que la medida clasifique correctamente a los pacientes que no mostraron cambios en el estándar externo). Es decir, se trata de valorar la capacidad de la medida para reflejar diferencias de cambio entre los grupos respecto al estándar externo (mejoría/ausencia de mejoría; empeoramiento/ausencia de empeoramiento).

El área bajo la curva (AUC) expresa la capacidad de discriminación del instrumento o probabilidad de clasificar correctamente tanto a los pacientes que mejoran (o

empeoran) como a los que no mejoran (o no empeoran). La curva ROC proporciona una visión general de la relación entre una medida y un estándar externo de cambio. Las principales desventajas son que la variable de clasificación (estándar) debe dicotomizarse, lo que puede hacer que se pierda información sobre la magnitud del cambio, y que requiere análisis independientes para definir sensibilidad al cambio para mejoría y para empeoramiento<sup>8,15</sup>.

### Diferencias en las medias de las puntuaciones de cambio

El estudio de las diferencias en las medias de las puntuaciones de cambio entre los diferentes subgrupos de la muestra puede realizarse mediante un análisis de la variancia (ANOVA) de medidas repetidas con un factor intrapacientes (ocasión de medida con dos niveles) y un factor entre pacientes (magnitud del cambio con 2 o más niveles: mejoría/sin mejoría). Los resultados del ANOVA permiten obtener la extensión con que los subgrupos difieren en cuanto al cambio, representada por la significación del término de interacción (grupo  $\times$  ocasión de medida).

### S de Norman

La S de Norman es un estadístico derivado de un análisis de la variancia de medidas repetidas y representa la relación entre la variancia de la interacción y la suma de ésta y la variancia del error<sup>16</sup>.

### Composición heterogénea: contraste entre pacientes

En este diseño, al igual que en el homogéneo, la muestra está formada por un único grupo. Sin embargo, los pacientes no son homogéneos entre ellos con cuanto al cambio, sino que es previsible que cambien de diferente manera. Un aspecto esencial es que se aplica un estándar externo cuyas puntuaciones de cambio son comparadas con las de la medida en cuestión. Por consiguiente, la sensibilidad al cambio dependerá de la elección del estándar, y no de las intervenciones realizadas. En estos casos lo que se evalúa es la sensibilidad al cambio externa. Los estadísticos más utilizados en estos diseños son el coeficiente de correlación y los modelos de regresión.

### Correlación

El coeficiente de correlación de Pearson se calcula por las puntuaciones de cambio entre dos medidas ( $x$  e  $y$ ).

$$r_{xy} = \frac{\sum_{i=1}^n (D_{xi} - \bar{D}_x)(D_{yi} - \bar{D}_y)}{n}$$

El coeficiente de correlación indica cómo cambian simultáneamente las dos medidas y sus valores oscilan entre  $-1$  y  $+1$ . En general,  $x$  es la medida en estudio e  $y$  es el resultado clínico específico (p. ej. un índice de capacidad funcional). Si  $r_{xy}$  se acerca a 1, la medida captura la información contenida en  $y$ , es decir, el instrumento responde a los cambios en el resultado clínico.

Las principales limitaciones del coeficiente de correlación derivan de su variación en respuesta a selección de valores concretos en una variable (p. ej., eliminación de valores extremos) y a la presunción de linealidad entre ambas variables, cuando en realidad esta relación puede no ser lineal<sup>8</sup>.

### Modelos de regresión

Un aspecto muy interesante de este diseño es examinar si los cambios en una medida son capaces de predecir cambios en la otra. Este análisis puede realizarse mediante modelos de regresión.

$$D_{yi} = \alpha + \beta d_{xi} + e_i$$

El parámetro  $\alpha$  representa el cambio medio en el resultado clínico ( $D_y$ ) en ausencia de cambio en el instrumento evaluado ( $d_x = 0$ ). El coeficiente  $\beta$  indica el incremento en el cambio del resultado clínico ( $D_y$ ) por cada unidad de cambio en el instrumento ( $d_x$ ). Valores de  $\beta$  cercanos a 0 indican que cambios importantes en  $x$  pueden no acompañarse de cambios en  $y$ , mientras que valores elevados de  $\beta$  indican que los cambios asociados en  $y$  también serán importantes. Este modelo puede generalizarse añadiendo más variables que predigan el cambio, incluso la situación clínica basal ( $Y_1$ ). Del mismo modo, es posible estandarizar los coeficientes para propósitos de comparación con otros estudios<sup>5</sup>.

Las principales limitaciones de este diseño se relacionan con la elección del estándar externo. A pesar de un coeficiente de correlación elevado, el estándar puede no capturar todos los cambios observados en los pacientes y por otra parte, el estándar y el instrumento pueden estar midiendo diferentes aspectos de un mismo concepto.

Para facilitar la comprensión de todos estos índices se presenta un ejemplo. Imaginemos que se realiza un estudio en el que se aplica una escala de salud, con un rango de puntuación de 0 a 25, a un grupo de 20 pacientes en una visita basal y al cabo de 6 meses de seguimiento. Durante la visita de seguimiento se pide al paciente una valoración general del cambio que ha experimentado sobre una escala de  $-5$  a  $5$ . Según esta valoración general, se crea una nueva variable, denominada grupo, que clasifica a los pacientes según hayan

experimentado o no un cambio (valores 1 y 0, respectivamente) utilizando un punto de corte = 4 (tabla 2).

Con estos datos, y sin hacer presunciones sobre el cambio esperable, se presentan diferentes tipos de análisis y estadísticos según las dimensiones de la sensibilidad al cambio (interna y externa) y las diversas características del cambio previsible (homogéneo, heterogéneo entre grupos, heterogéneo entre pacientes).

Si asumimos que la muestra está formada por un único grupo de pacientes que previsiblemente cambiarán más o menos de la misma forma, el estadístico que hay que calcular es la SRM o cociente entre el cambio medio y su DE. En nuestro ejemplo se obtendrá una SRM = 0,69 (3,4 / 4,9); es decir, la medida en cuestión tendría una sensibilidad al cambio moderada. Este estadístico no informa de posibles diferencias en el cambio entre las dos categorías definidas por la variable grupo.

Supongamos, por el contrario, que la muestra está formada por dos subgrupos de pacientes en los que el cam-

bio se produce de distinta forma. En el ejemplo anterior, estos dos grupos estarían representados por los valores 0 y 1 de la variable grupo. En este caso, la sensibilidad al cambio puede estudiarse mediante la construcción de curvas ROC entre la diferencia de las puntuaciones y la categoría de cambio (grupo). Los resultados ofrecen un AUC de 0,869 (0,715-1), con unos valores de sensibilidad y especificidad del 78 y el 73%, respectivamente, para una diferencia entre la puntuación basal y la de seguimiento = 5 unidades (tabla 3). La principal limitación de este procedimiento puede estar en relación con la necesidad de dicotomizar los resultados del estándar externo. De esta forma sólo se han considerado las categorías "mejoría/sin mejoría", aunque ha habido pacientes que han podido empeorar y para los que habría que haber definido una nueva clasificación.

Otra forma de abordar el contraste entre subgrupos de la muestra es realizar un ANOVA de medidas repetidas con un factor intrapaciente (ocasión de medida) y otro entre pacientes (tipo de cambio). El término de interacción ocasión de medida-tipo de cambio informa de la extensión con la que los dos grupos cambian de diferente manera. En nuestro ejemplo, el término de interacción es significativo, por lo que la magnitud del cambio es diferente entre ambos grupos (tabla 4).

Imaginemos finalmente que nuestra muestra está formada por un grupo heterogéneo de individuos que previsiblemente cambiarán de diferente forma. En esta situación puede utilizarse un análisis de correlación entre el cambio producido en la medida en cuestión y en un estándar externo. En nuestro ejemplo el coeficiente de correlación de Spearman entre el cambio producido (diferencia entre las puntuaciones basal y seguimiento) y el estándar externo (valoración general del paciente) es de 0,71.

**TABLA 2. Datos de un estudio hipotético**

Paciente	Basal	Seguimiento	Diferencia	Valoración general	Grupo
1	25	18	7	5	1
2	20	20	0	-3	0
3	15	6	9	4	1
4	9	5	4	5	1
5	24	12	12	4	1
6	15	18	-3	-4	0
7	8	6	2	3	0
8	12	6	6	4	1
9	15	10	5	4	1
10	14	7	7	5	1
11	12	12	0	-3	0
12	8	5	3	5	0
13	12	8	4	1	0
14	20	15	5	2	1
15	10	19	-9	-4	0
16	12	5	7	4	1
17	17	10	7	3	1
18	9	3	6	3	0
19	21	21	0	-4	0
20	18	22	-4	0	0
Media	14,8	11,4	3,4	1,7	
Desviación estándar	5,2	6,3	4,9	3,4	

**TABLA 3. Curva de rendimiento diagnóstico**

Punto de corte	Sensibilidad (%)	Especificidad (%)	Clasificación correcta (%)	Razón de verosimilitud +
≥ -9	100	0	45	1,0
≥ -4	100	9,1	50	1,1
≥ -3	100	18,2	55	1,2
≥ 0	100	27,3	60	1,4
≥ 2	100	54,5	75	2,2
≥ 3	100	63,6	80	2,7
≥ 4	88,9	63,6	75	2,8
≥ 5	77,8	72,7	75	2,8
≥ 6	66,7	81,8	75	3,7
≥ 7	55,6	90,9	75	6,1

TABLA 4. ANOVA de medidas repetidas

Origen	Suma cuadrados	gl	Media cuadrática	F (p)
Entre sujetos	2,5	1	2,5	
Cambio	2,5	1	2,5	0,04 (0,837)
Intrasujetos	238,6	2	119,05	
Visita	115,6	1	115,6	19,11 (0,0004)
Visita × cambio	122,5	1	122,5	20,25 (0,0003)
Residual	108,9	18	6,0	

En los estudios de sensibilidad al cambio es importante tener en cuenta las dos dimensiones, interna y externa, y la utilización de un diseño adecuado que permita evaluar posibles diferencias de cambio entre grupos o pacientes<sup>2,8,17</sup>.

En este sentido, Veehof et al han publicado recientemente un estudio sobre las propiedades psicométricas de dos índices de actividad de la artritis reumatoide (RADAI y su forma abreviada) en una cohorte de pacientes con esta enfermedad que inician tratamiento contra el factor de necrosis tumoral (TNF). Los autores plantean el estudio de la sensibilidad al cambio para la disminución de la actividad (mejoría) subdividiéndola en sus dos dimensiones, interna y externa. En el caso de la interna, se calcula la SMR junto con su intervalo de confianza. Para la externa se utiliza como estándar externo los criterios EULAR de respuesta que clasifican a los pacientes en respondedores y no respondedores.

Las diferencias de cambio entre ambos grupos no son capturadas por la SMR, ya que quedan contenidas en su variabilidad. Por consiguiente, los autores utilizan curvas ROC entre las puntuaciones de cambio y el criterio externo para evaluar la capacidad de discriminación entre ambos grupos. Además, se calcula la SMR dentro de cada uno de ellos.

Los resultados muestran una sensibilidad al cambio interna y externa moderadas (SMR, 0,76 y 0,80; AUC, 0,77 y 0,78). Los respondedores tuvieron mejorías importantes en la actividad de la enfermedad (SMR > 0,80), mientras que no se objetivó mejoría en los no respondedores (SMR < 0,20)<sup>18</sup>.

### Limitaciones relacionadas con el uso de coeficientes inadecuados

Los diferentes métodos de medición de la sensibilidad al cambio tienen distintos objetivos, por lo que conducen a diferentes conclusiones<sup>2</sup>. La utilización de un coeficiente inadecuado para el tipo de diseño en estudio puede hacer que la señal (cambio real) de algunos coeficientes esté incluida en el ruido (variabilidad) de otros, por lo que puede pasar inadvertida. Por ejemplo, si se evalúa la sensibilidad al cambio con una SMR so-

bre una muestra formada por subgrupos o pacientes que cambiarán de diferente forma, las puntuaciones de cambio entre grupos o entre pacientes estarán contenidas en la porción de variabilidad (ruido) de la SMR, lo que disminuirá su magnitud. Sin embargo, a pesar de que la señal del cambio entre grupos o entre pacientes pueda estar contenida en el ruido de la SMR, es posible obtener una SMR distinta de cero por diferentes razones<sup>8</sup>.

En primer lugar, además de la estimación puntual, debe calcularse un intervalo de confianza para el coeficiente utilizado con el fin de evaluar la probabilidad de que la estimación sea realmente diferente de cero<sup>19</sup>.

En segundo lugar, los investigadores interesados en evaluar la capacidad de una medida para detectar cambios seleccionan a pacientes que en general se espera que mejoren. Por lo tanto, el cambio medio de la muestra será mayor que cero aun cuando algunos pacientes permanezcan estables o incluso empeoren. Cuando el cambio medio es mayor que cero, la SMR será mayor que cero incluso aunque existan subgrupos o pacientes individuales que cambian de diferente manera.

En tercer lugar, puede ocurrir que en una muestra realmente homogénea con respecto al cambio se observen diferencias aparentes en el cambio entre pacientes, representadas por una elevada correlación con otra medida. En general, estos casos se deben a una presunción sobre la existencia de correlación entre las puntuaciones de cambio en la medida y la valoración retrospectiva general del propio paciente. Sin embargo, la utilización de una valoración retrospectiva sobrestima la correlación existente entre ambas medidas. Para entender el mecanismo de esta asociación aparente, es preciso considerar las relaciones y las presunciones sobre las puntuaciones observadas (suministradas por los pacientes), reales (valores desconocidos que representan las puntuaciones que se obtendrían en ausencia de error de medición) y los errores de medición (diferencia entre las puntuaciones reales y las observadas). En teoría, las puntuaciones de cambio y la valoración retrospectiva están midiendo un mismo atributo, pero se considera que ambas medidas son independientes y, por lo tanto, los errores de medición no están correlacionados. Sin embargo, es muy poco probable que los errores sean independientes cuando es el propio paciente el que aporta las puntuaciones de cambio tanto en la medida como en la valoración general. La consecuencia es que la correlación observada será mayor que cero incluso cuando la correlación entre las puntuaciones reales sea cero<sup>4</sup>. En este sentido, Fransen et al<sup>20</sup> compararon la sensibilidad al cambio de medidas basadas en la percepción del paciente y medidas objetivas para detectar exacerbaciones de la artritis reumatoide. Los resultados demostraron menor sensibilidad al cambio de los ítems de naturaleza subjetiva. Por consiguiente, aunque los coeficientes de sensibilidad al

cambio puedan ser similares, las medidas de desenlace subjetivas y objetivas no son intercambiables.

Por consiguiente, y como conclusión, antes de realizar un estudio de sensibilidad al cambio es importante conocer la validez y la reproducibilidad del instrumento en cuestión. Además es fundamental realizar un buen diseño de la muestra de análisis, definir las características del cambio previsible y elegir un estándar externo adecuado en los casos necesarios. Es importante tener en cuenta que la aplicación de medidas de sensibilidad al cambio inadecuadas puede originar resultados poco fiables.

## Bibliografía

- Guyatt GH, Kirshner B, Jaeschke R. Measuring health status: what are the necessary measurement properties? *J Clin Epidemiol.* 1992;45:1341-5.
- Terwee CB, Dekker FW, Wiersinga WM, Prummel MF, Bossuyt PM. On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. *Qual Life Res.* 2003;12:349-62.
- Beaton DE, Hogg-Johnson S, Bombardier C. Evaluating changes in health status: reliability and responsiveness of five generic health status measures in workers with musculoskeletal disorders. *J Clin Epidemiol.* 1997;50:79-93.
- Norman GR, Stratford P, Regehr G. Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. *J Clin Epidemiol.* 1997;50:869-79.
- Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol.* 2000;53:459-68.
- Walsh TL, Hanscom B, Lurie JD, Weinstein JN. Is a condition-specific instrument for patients with low back pain/leg symptoms really necessary? The responsiveness of the Oswestry Disability Index, MODEMS, and the SF-36. *Spine.* 2003;28:607-15.
- De Vet HC, Bouter LM, Bezemer PD, Beurskens AJ. Reproducibility and responsiveness of evaluative outcome measures. Theoretical considerations illustrated by an empirical example. *Int J Technol Assess Health Care.* 2001;17:479-87.
- Stratford PW, Riddle DL. Assessing sensitivity to change: choosing the appropriate change coefficient. *Health Qual Life Outcomes.* 2005;3:23.
- Koh ET, Leong KP, Tsou IY, et al. The reliability, validity and sensitivity to change of the Chinese version of SF-36 in oriental patients with rheumatoid arthritis. *Rheumatology (Oxford).* 2006;45:1023-8.
- Beaton DE, Bombardier C, Katz JN, Wright JG. A taxonomy for responsiveness. *J Clin Epidemiol.* 2001;54:1204-17.
- Corzillius M, Fortin P, Stucki G. Responsiveness and sensitivity to change of SLE disease activity measures. *Lupus.* 1999;8:655-9.
- Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. *Med Care.* 1990;28:632-42.
- Beaton DE, Boers M, Wells GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. *Curr Opin Rheumatol.* 2002;14:109-14.
- Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. *J Chronic Dis.* 1986;39:897-906.
- Stratford PW, Binkley FM, Riddle DL. Health status measures: strategies and analytic methods for assessing change scores. *Phys Ther.* 1996;76:1109-23.
- Norman GR. Issues in the use of change scores in randomized trials. *J Clin Epidemiol.* 1989;42:1097-105.
- Salaffi F, Stancati A, Neri R, Grassi W, Bombardieri S. Measuring functional disability in early rheumatoid arthritis: the validity, reliability and responsiveness of the Recent-Onset Arthritis Disability (ROAD) index. *Clin Exp Rheumatol.* 2005;23 Suppl 39:S31-42.
- Veehof MM, Ten Klooster PM, Taal E, Van Riel PL, Van de Laar MA. Psychometric properties of the Rheumatoid Arthritis Disease Activity Index (RADAI) in a cohort of consecutive Dutch patients with RA starting anti-tumour necrosis factor treatment. *Ann Rheum Dis.* 2008;67:789-93.
- Wright JG, Young NL. A comparison of different indices of responsiveness. *J Clin Epidemiol.* 1997;50:239-46.
- Fransen J, Hauselmann H, Michel BA, Caravatti M, Stucki G. Responsiveness of the self-assessed rheumatoid arthritis disease activity index to a flare of disease activity. *Arthritis Rheum.* 2001;44:53-60.