

CUESTIONARIOS EN ATENCIÓN PRIMARIA

Cuestionarios de salud: concepto y metodología

L. Prieto^{a,b,c} y X. Badia^a

^aUnitat de Recerca en Resultats en Salut. Servei d'Epidemiologia Clínica i Salut Pública. Hospital de la Santa Creu i Sant Pau. Barcelona. ^bFacultat de Psicologia i Ciències de l'Educació Blanquerna. Universitat Ramon Llull. Barcelona. ^cClassification, Assessment, Surveys and Terminology (CAS). World Health Organization. Ginebra. Suïça.

Introducción

En la actualidad se reconoce de forma generalizada que el impacto que las enfermedades tienen sobre los pacientes no puede ser descrito en su totalidad por medidas objetivas de la salud como pueden ser la extensión de un tumor, el resultado de una biopsia o la medición de la tensión arterial. La consideración de otros factores, que podríamos calificar como subjetivos, tales como el dolor, la capacidad funcional o el bienestar emocional, se consideran imprescindibles a la hora de evaluar de forma global el estado de salud de las personas¹. De este reconocimiento surge un área de investigación conocida como «calidad de vida relacionada con la salud» (CVRS).

La evaluación CVRS se ha convertido en un importante objetivo a tener en cuenta en el transcurso de cualquier investigación clínica. A pesar de ello, el término CVRS se utiliza a menudo de forma vaga y sin una definición clara. Esto no resulta sorprendente, dada la amplia naturaleza de un concepto que incluye, entre otros, aspectos relacionados con la función física (entendida como la capacidad para llevar a cabo las actividades más comunes de la vida

diaria, tales como el autocuidado o el desplazamiento), la función social (relaciones con los demás y participación en actividades sociales), la percepción del estado de salud general, el dolor o la satisfacción con la vida.

En los últimos 20 años se han desarrollado infinidad de instrumentos en forma de cuestionario destinados a la medición de la CVRS^{2,3}. Algunos de estos instrumentos (p. ej., EQ-5D, SF-36, SIP, NHP) pueden considerarse genéricos³, ya que evalúan de forma general aspectos relacionados con la salud física, mental y social, y pueden ser utilizados por tanto para comparar enfermedades y tratamientos. Otros instrumentos, definidos como específicos^{3,4}, diseñados para medir la CVRS en una enfermedad o grupo de enfermedades concreto, complementan e incluso sustituyen los resultados proporcionados por los cuestionarios genéricos. En cualquier caso, y si bien esta dicotomía puede ser útil para la clasificación de los instrumentos de evaluación de la CVRS, no hay que olvidar que existen otros instrumentos o cuestionarios de salud que no necesariamente entran dentro de esta definición; la oferta es diversa⁴: existen instrumentos diseñados para evaluar exclusivamente un único aspecto relacionado con la CVRS (p. ej., dolor, depresión) o instrumentos en los que el propio paciente identifica y valora los aspectos más importantes para definir su calidad de vida⁵. Existen referencias bibliográficas que dan buena cuenta de los instrumentos

existentes, su contenido, sus características y su propósito^{3,6}.

Estos cuestionarios de evaluación de la CVRS a los que nos referimos se utilizan cada vez más en ensayos clínicos multicéntricos, estudios epidemiológicos o en la práctica clínica cotidiana^{7,8}. Las ventajas que se derivan de la evaluación estandarizada de los aspectos más subjetivos relacionados con la salud son diversas, pero también los inconvenientes ligados a la medición de conceptos intangibles, latentes y no observables en todos los casos. La necesidad de una adecuada interpretación de los resultados proporcionados por dichos instrumentos, así como la capacidad de los usuarios potenciales para seleccionar el instrumento que optimice sus intereses es ahora más grande que nunca. En este sentido, hay que resaltar que el correcto uso de los instrumentos de medición de la CVRS pasa necesariamente por la comprensión del proceso metodológico subyacente a su construcción y funcionamiento.

El objetivo del presente artículo es, por tanto, introducir al profesional de la atención primaria (AP) lego en la materia en el conocimiento de los aspectos conceptuales y metodológicos básicos que determinan el funcionamiento de los cuestionarios de evaluación de la CVRS.

El proceso psicométrico

Los cuestionarios de medida de la CVRS proporcionan invariablemente una puntuación (o puntuación

Correspondencia: Dr. Luis Prieto.
World Health Organization. Classification,
Assessment and Survey Group (CAS). Global
Programme on Evidence for Health Policy (GPE).
Avenue Appia 20. CH-1211 Geneva 27.
Switzerland.
Correo electrónico: prietol@who.int

(Aten Primaria 2001; 28: 201-209)

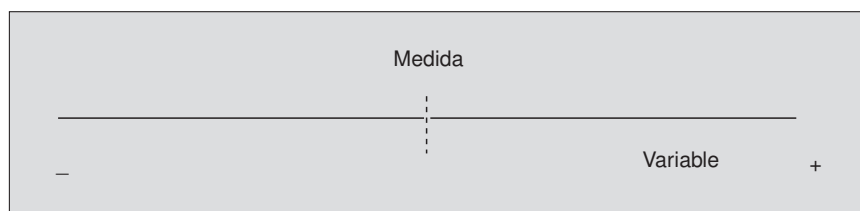


Figura 1. La idea de medida.

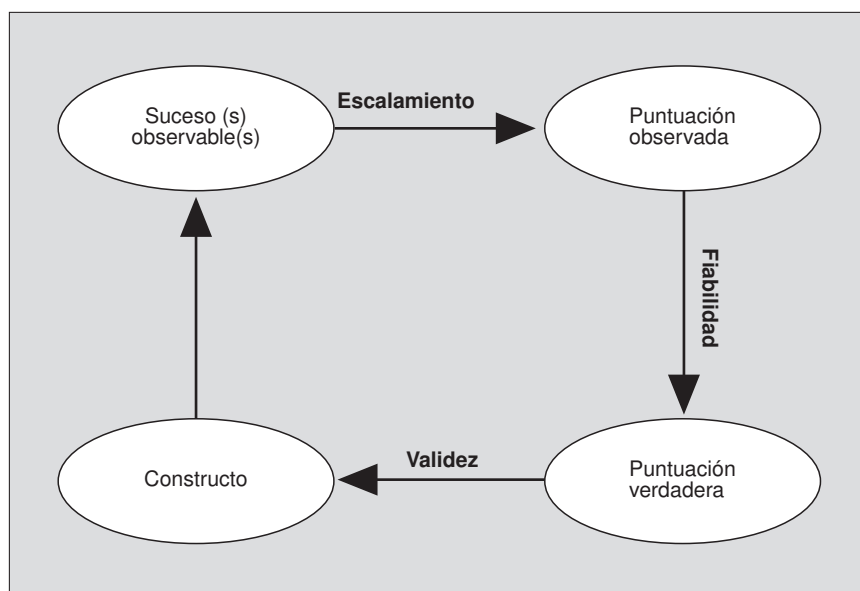


Figura 2. El proceso psicométrico.

ciones) para cada uno de los pacientes a los que se aplican. Esta puntuación, expresada en unas unidades siempre arbitrarias (p. ej., de 0 a 100 puntos), pretende ser una medida de aquella característica que está siendo evaluada.

La idea, abstracta sin duda, de qué es una medida requiere la idea, también abstracta, de una variable en la cual situar dicha medida⁹. Si la variable se visualiza como una línea, la medida puede ser representada como un punto sobre esa misma línea. Esta relación entre la idea de medida y su variable se representa en la figura 1.

Tómese como ejemplo de variable el dolor. Por propia experiencia, se sabe que el dolor puede variar en intensidad. La línea de la figura 1

representa la idea de dolor y su intensidad: cuanto más a la derecha de la línea, más dolor, cuanto más a la izquierda, menos dolor. Cuando administramos a un paciente un cuestionario que evalúa el dolor que presenta, nuestra intención fundamental es estimar su posición sobre la línea implicada por dicho cuestionario.

Para poder hacer esto, en la práctica deberemos disponer de un cuestionario que defina la línea, así como de un sistema que nos permita transformar las respuestas de los pacientes en el cuestionario en un punto sobre ella^{9,10}. Solventado este primer problema, se plantean dos cuestiones adicionales relacionadas con el establecimiento de cualquier medida, sea cual sea su naturaleza: en primer lugar, ¿es fiable la estimación ob-

tenida?, o dicho de otro modo, ¿dadas las mismas condiciones, obtendrá el paciente la misma puntuación/posición sobre la línea?, y en segundo lugar, ¿es válida la estimación obtenida?, es decir, ¿es realmente una medida del dolor que presenta esa persona o es fruto de otra variable diferente (p. ej., la ansiedad que presente el paciente en cuestión)?

El proceso de medición subyacente a la descripción que acabamos de realizar se representa de forma esquemática en la figura 2. Este proceso, o manera de operar, se denomina habitualmente «proceso psicométrico»¹¹. La razón para emplear dicha expresión es simple: los fundamentos de la medición de la salud a través de cuestionarios no son más que una herencia de la metodología utilizada en psicología para la medición de otros fenómenos psicológicos (de ahí el nombre de la disciplina dedicada a tal fin: psicometría), tales como la inteligencia o la personalidad, que al igual que los fenómenos implicados en la definición de la salud tienen un carácter totalmente subjetivo⁹⁻¹¹.

A continuación se verá con más detalle cuáles son los elementos del esquema propuesto.

Constructo

Como ya se ha apuntado anteriormente, el objeto de cualquier medición es siempre una variable⁹. Cuando dicha variable es abstracta y latente, en lugar de concreta y observable, ésta se denomina «constructo»¹⁰. Tal variable es, de forma literal, «algo» artificialmente «construido», «algo» definido de forma teórica por su posible utilidad práctica. Un constructo es un concepto hipotético cuya existencia nunca podrá ser absolutamente confirmada.

El establecimiento de cualquier teoría implica la conceptualización de una serie de constructos, el desarrollo de medidas de esos constructos y la determinación de

las relaciones funcionales entre las medidas obtenidas⁹⁻¹¹.

Los diferentes aspectos incluidos en la definición de CVRS son precisamente constructos: la función física, la función social o el dolor antes mencionados no son más que representaciones teóricas de conceptos intangibles e inobservables directamente. La metodología que nos ocupa nos permitirá desarrollar medidas específicas para esos constructos, que a su vez nos permitirán responder a preguntas que implican a varios de esos constructos (p. ej., ¿en qué medida se ven afectadas las actividades sociales de una persona cuando presenta dolor?).

Suceso(s) observable(s)

Antes de que se pueda llevar a cabo la medición de un constructo, es necesario establecer alguna regla de correspondencia entre dicho constructo teórico y ciertos sucesos observables que resulten ser legítimos indicadores de éste¹¹. Para medir el «dolor» de un paciente, el investigador tendrá primero que especificar cuáles son los sucesos observables que indican la presencia o ausencia de tal concepto, tendrá por tanto que establecer una definición operativa del dolor.

En el área que nos ocupa, la definición operativa tomará forma de preguntas o afirmaciones en un cuestionario estandarizado en el que todos los pacientes deberán responder de la misma forma. Así, por ejemplo, las respuestas positivas (p. ej., ante dos opciones de respuesta posibles: «Sí» y «No») de un paciente a afirmaciones tales como: «Tengo dolor por las noches», «Tengo dolor al subir y bajar escaleras» o «Tengo dolor al cambiar de postura» supondrán una confirmación de la presencia de cierta magnitud del constructo «dolor» en dicho paciente.

El contenido de los cuestionarios de CVRS supone la definición operativa de aquello que preten-

den medir. Las respuestas de los pacientes a las preguntas o afirmaciones que éstos formulan representan la concreción observable del constructo que está siendo evaluado¹². Los posibles sucesos observables que definen el constructo son prácticamente infinitos, por lo que podemos considerar que el contenido final del cuestionario no es más que una muestra representativa de todos aquellos sucesos (léase preguntas o afirmaciones) que podrían definir el constructo de interés. Por esta razón, es habitual encontrar que el contenido de diferentes cuestionarios que pretender evaluar el mismo concepto no es idéntico.

Con el ánimo de estandarizar al máximo la medición, las respuestas a las preguntas de los cuestionarios de CVRS se encuentran habitualmente limitadas y con un formato fijo. Si bien el formato puede ser muy diverso¹³, las opciones de respuesta dicotómicas (p. ej., sí/no; verdadero/falso...) y politómicas (p. ej., siempre/frecuentemente/a veces/raras veces/nunca; ningún dolor/muy poco/un poco/moderado/mucho/muchísimo...) son las más frecuentes.

Escalamiento y puntuación observada

Dado que el cuestionario contiene los sucesos observables necesarios para la definición del constructo a medir, el siguiente paso en el proceso psicométrico consistirá en la determinación de un sistema que permita transformar las respuestas de los pacientes al cuestionario en un punto sobre la línea imaginaria que éste define. El «sistema» o procedimiento a seguir para alcanzar este propósito se denomina escalamiento (scaling)⁹⁻¹¹.

Cualquier procedimiento de escalamiento implicará el establecimiento de un vínculo de correspondencia entre los elementos que están siendo evaluados por el

cuestionario (los pacientes) y el sistema de números reales. Recuérdese que el sistema de números reales está comprendido por el 0 y todos los demás valores numéricos, tanto enteros como decimales, situados entre menos infinito y más infinito¹⁰. No es una coincidencia que el sistema de números reales pueda ser representado gráficamente mediante una línea que, a su vez, puede ser infinitamente dividida en segmentos más y más pequeños. Cada valor en el sistema de números reales tiene una única localización (un punto) sobre esa línea.

Imagínese que se tiene una serie de varillas que claramente difieren en longitud. Fácilmente se podrían establecer diferentes procedimientos para asignar un valor numérico real a cada uno de ellos. Por ejemplo, se podrían ordenar las varillas de menor a mayor longitud y asignar un número a cada una de ellas en función de su posición: a la más corta se le puede asignar el número 1, a la siguiente más larga el número 2 y así sucesivamente. Podríamos también fácilmente comparar las varillas con una unidad estándar como es el centímetro y asignar a cada varilla números basados en su longitud «en centímetros». En ambos casos se habría completado la tarea de escalamiento con las varillas y obtenido una «puntuación observada» para cada una de ellas.

Si bien la tarea del escalamiento puede parecer obvia cuando se manejan conceptos que se podrían definir como «del mundo físico», no lo es tanto cuando nos enfrentamos con constructos abstractos como los que definen la CVRS.

La determinación de una unidad estándar de comparación para convertir las respuestas de los pacientes a un cuestionario en una puntuación observada resulta prácticamente imposible. Por esta razón, se han tenido que plantear estrategias de escalamiento alternativas fundamentadas en teo-

TABLA 1. Respuestas de 10 pacientes (P1-P10) a 4 ítems (I1-I4) sobre el dolor

	I1	I2	I3	I4	Suma	Puntuación 0-100
P1	1	1	1	1	4	100
P2	0	1	1	1	3	75
P3	1	1	1	1	4	100
P4	1	1	0	1	3	75
P5	0	0	0	0	0	0
P6	1	0	1	0	2	50
P7	1	1	0	0	2	50
P8	1	0	1	1	3	75
P9	1	1	1	1	4	100
P10	0	1	0	0	1	25

Ítems					
I1.	Tengo dolor por las noches	Sí	No		
I2.	Tengo dolor al subir y bajar escaleras	Sí	No		
I3.	Tengo dolor al cambiar de postura	Sí	No		
I4.	Tengo dolor al caminar	Sí	No		

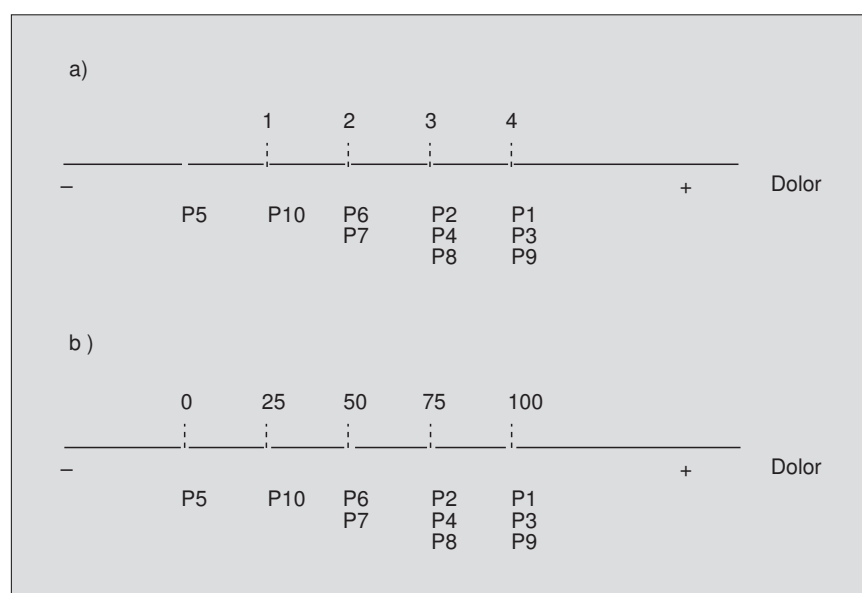


Figura 3. Localización de los 10 pacientes (P1-P10) de la tabla 1 sobre la línea de medida implicada por los 4 ítems de dolor.

rías matemáticas de mayor o menor complejidad. La teoría conocida como «teoría clásica de los tests» (TCT)¹⁰ aporta una solución para la obtención de la puntuación observada en un cuestionario, que si bien puede parecer simple resuelve de forma elegante el problema planteado: la suma de las respuestas otorgadas a los ítems que componen el cuestionario resulta ser una estimación

bastante aproximada del «punto» que el paciente ocupa sobre la línea que define dicho cuestionario. La tabla 1 muestra un ejemplo hipotético con las respuestas proporcionadas por 10 pacientes (P1-P10) a un cuestionario de dolor que contiene 4 afirmaciones con opciones de respuesta dicotómicas: «Sí/No»: «Tengo dolor por las noches», «Tengo dolor al subir y bajar escaleras», «Tengo dolor al

cambiar de postura», «Tengo dolor al caminar».

El primer paso en la obtención de una puntuación observada para cada uno de los pacientes consiste en asignar un valor numérico a cada una de las posibles respuestas. Suponiendo que las respuestas, afirmativas o negativas, pueden considerarse como sucesos observables que confirman la presencia/ausencia de dolor, asignar los números 1 al «Sí» y 0 al «No» parece razonable.

En la tabla 1 se muestra, por ejemplo, que el paciente 1 (P1) ha respondido con un «Sí» a las 4 afirmaciones, mientras que el paciente 5 ha respondido con un «No» a todas ellas. La suma de las puntuaciones se muestra en la columna 6. La figura 3a representa la posición que cada uno de los pacientes ocuparía sobre la línea que representa la variable evaluada. Es importante destacar que los números utilizados para representar las posiciones de las personas son totalmente arbitrarios y relativos (el valor 0 no representa en este caso la ausencia «absoluta» de dolor): la puntuación observada obtenida puede ser transformada de forma lineal sin que se produzca ninguna diferencia sustantiva en los resultados obtenidos, aparte claro está del cambio en las unidades de medida (p. ej., la suma de las puntuaciones de la tabla 1 ha sido transformada a una escala de 0-100 puntos: véase tabla 1, columna 7 y fig. 3b). Los valores asignados a las opciones de respuesta también podrían ser modificados sin mayores consecuencias (el lector puede comprobar por sí mismo que la posición de las personas sobre la línea resulta invariante al sustituir los valores 1 [sí] y 0 [no] por cualquier otro par de números que mantengan las mismas propiedades ordinales [$1 > 0$]).

La mayoría de los cuestionarios de medida de la CVRS utilizan la TCT como base para el escala-

miento de sus respuestas y para la obtención de una puntuación observada. Paradójicamente, los supuestos que subyacen a la TCT han sido muy discutidos, especialmente aquellos referentes a la adecuación de la construcción de escalas continuas de medida (como la definida en la figura 3), a partir de variables de naturaleza claramente categórica (p. ej., las opciones de respuestas «Sí/No»). Diversos modelos de escalamiento han sido planteados como solución alternativa a la TCT. Entre estos modelos destaca el propuesto por el matemático danés Georg Rasch, quien en 1960¹⁴ definió una nueva metodología para escalar las respuestas a los ítems de un cuestionario a partir del uso de la función logística. Este nuevo modelo, el modelo de Rasch, tiene sus fundamentos en una teoría psicométrica conocida como «teoría de la respuesta al ítem» (TRI)¹⁵. Sus defensores indican que los supuestos de la TRI permiten corregir los problemas relacionados con la continuidad de las escalas definidas a partir de datos categóricos. En la actualidad son ya diversos los instrumentos de CVRS que, de una u otra manera, empiezan a incorporar esta nueva metodología^{4,16-19}.

Antes de finalizar este apartado, es importante destacar que, independientemente de la técnica de escalamiento empleada, los cuestionarios de CVRS pueden proporcionar una única puntuación observada (se les suele denominar «índices» en tal caso), o también diversas puntuaciones, una para cada uno de los diferentes constructos evaluados por el instrumento (se habla entonces de «perfiles», dado que las diferentes puntuaciones obtenidas proporcionan algo así como un perfil del estado de salud del paciente)^{3,6}. En algunas ocasiones los perfiles pueden convertirse también en una única puntuación mediante algún algoritmo que permita combinar los diferentes valores nu-

méricos en un único índice. La decisión que llevará a los autores de un cuestionario a proponer una puntuación observada con formato de índice, perfil o ambos, es arbitraria y se basará en la naturaleza atribuida al objeto de medida del instrumento²⁰. La decisión puede ser polémica, ya que diferentes autores pueden considerar un mismo objeto como uni o multidimensional, dependiendo del punto de vista empleado. Así, para algunos autores es legítimo proporcionar una única puntuación para el concepto «salud», mientras que otros consideran necesario desgranar los resultados en diferentes puntuaciones, una para cada «constructo» subyacente a la idea de «salud».

La discusión surge precisamente ante la reconocida necesidad de que toda medida sea unidimensional. Es imposible hacer una definición de la línea de medida antes descrita en más de una dimensión. Es obvio que se puede medir el peso y la altura de un individuo y combinar la información para obtener una nueva medida que pueda ser útil en la práctica (p. ej., el índice de masa corporal [IMC]), pero también es obvio que no se puede medir «de una sola vez» el IMC, precisamente por su carácter bidimensional. A pesar de que existen técnicas (p. ej., análisis factorial²¹) que permiten una aproximación empírica a la respuesta más adecuada: unidimensionalidad o multidimensionalidad de los sucesos observables incluidos en un cuestionario^{22,23}, la decisión final no es siempre fácil de tomar.

Puntuación verdadera

La esencia del proceso psicométrico que se está describiendo radica en el supuesto de que cualquier puntuación observada en un cuestionario es el resultado de combinar dos componentes hipotéticos: una puntuación verdadera y un error aleatorio⁹⁻¹¹. La idea se pue-

de expresar mediante la siguiente formulación:

$$X = V \pm E$$

en donde X representa la puntuación observada en el cuestionario; V, la puntuación verdadera, y E, el error aleatorio.

La idea es común a cualquier otro proceso de medida o evaluación. Supóngase que un médico sospecha que su paciente presenta un trastorno en el hígado. Para confirmar o descartar su hipótesis, el médico ordena una prueba de laboratorio (p. ej., un análisis enzimático). Se considerará el resultado de la prueba como una medición de la variable «presentar o no presentar el trastorno». Cualquiera de los dos valores, positivo (1 = lo presenta) o negativo (0 = no lo presenta), que dicha variable puede tomar será la «puntuación observada» de esa medida para ese paciente. Es obvio que para la variable biológica implicada en ese caso el paciente sí tiene una puntuación verdadera absoluta: positivo (1) o negativo (0). A pesar de que existe tal puntuación verdadera para el paciente, es posible que se produzca un error en el método de medida de dicha variable. Este error, seguramente pequeño, obedecerá a una multiplicidad de causas difíciles de predecir y controlar, de ahí que se defina como error aleatorio. Debido al error aleatorio, es posible que la prueba de laboratorio utilizada para detectar el trastorno pueda producir resultados incluso diferentes cuando ésta se repite en el mismo paciente. En cualquier caso, lo que más interesa resaltar es que, sea cual sea el resultado de la prueba, siempre existirá una puntuación verdadera absoluta, totalmente independiente del resultado de la misma. A diferencia de lo que ocurre en el caso de variables físicas o biológicas, la puntuación verdadera en un cuestionario de CVRS es un concepto estadístico y no una puntua-

ción realmente verdadera en el sentido más absoluto de la definición¹⁰. Imagínese que es posible administrar un cuestionario a un paciente infinitas veces de forma consecutiva, evitando de algún modo que en cada nueva ocasión recuerde las respuestas ofrecidas en la ocasión anterior. Obviamente las puntuaciones observadas obtenidas para cada una de esas repeticiones fluctuarán debido a errores, que consideraremos aleatorios como en el caso del primer ejemplo. Dada la aleatoriedad de los errores, las puntuaciones observadas se distribuirán de forma normal. La media de esa distribución será sin duda la puntuación que mejor representa el estado del paciente. Por tanto, y tratando de evitar cualquier formulación más críptica, se puede decir que la puntuación verdadera para un paciente en un cuestionario de CVRS se interpreta como el promedio de todas sus puntuaciones observadas a lo largo de una serie infinita de administraciones repetidas del mismo cuestionario¹⁰. Dada la abstracción del término, es evidente que la puntuación verdadera de una persona en un cuestionario de CVRS no podrá llegar a ser determinada nunca de forma empírica. En cualquier caso, la consideración de tal concepto (así como el de error aleatorio) será útil para establecer las estrategias necesarias para determinar una de las propiedades fundamentales de cualquier instrumento de medida: la fiabilidad.

Fiabilidad

Por lo anteriormente expuesto, está claro que el usuario de cualquier cuestionario de CVRS sólo puede conocer las puntuaciones observadas para el mismo, a pesar de que seguramente estaría más interesado en conocer las puntuaciones verdaderas para el mismo. Surge así una importante pregunta: ¿cuán relacionadas están

las puntuaciones observadas obtenidas para un grupo de pacientes con lo que serían sus puntuaciones verdaderas? Cuanta mayor sea la discrepancia, mayor será el error aleatorio presente en la observación realizada, y viceversa, cuanto más se parezcan la puntuación observada X y la puntuación verdadera V , menor será el error (E) contenido por X (recuérdese que se ha definido que $X = V \pm E$)^{9,10}.

Un posible índice de la magnitud de esta asociación sería la correlación entre ambas puntuaciones. De ahí que el coeficiente de correlación que expresa el grado de asociación entre las puntuaciones observadas y las puntuaciones verdaderas se conozca como índice de fiabilidad (ρ_{XV})¹⁰. Desgraciadamente esta expresión sigue sin tener gran utilidad: las puntuaciones verdaderas no son directamente observables y no podemos tener a los pacientes respondiendo infinitamente al cuestionario para disponer de todas las puntuaciones observadas.

Sin embargo, resulta plausible imaginar la administración repetida del cuestionario a un grupo de pacientes en dos ocasiones diferentes. Cuando el cuestionario cumple unos requisitos matemáticos descritos por la antes mencionada TCT^{9,10}, es posible establecer una correspondencia directa entre ρ_{XV} , la correlación entre puntuaciones observadas y verdaderas, y ρ_{XX} , la correlación entre las puntuaciones observadas en 2 administraciones paralelas, o dicho de otro modo, entre el test y el retest del instrumento en cuestión. Los requisitos matemáticos asumidos por la TCT implican que cada paciente tendrá la misma puntuación verdadera para cada una de las administraciones paralelas del cuestionario y que la variancia del error aleatorio sea también equivalente en ambas administraciones. Establecidos estos supuestos, se define un indicador, conocido como coefi-

ciente de fiabilidad (CF), a partir de la correlación entre las puntuaciones observadas en las 2 administraciones paralelas (test y retest) del instrumento^{9,10}.

El CF puede oscilar entre 0 y 1: cuanto mayor es su valor, mayor será la asociación entre las puntuaciones observadas y las verdaderas, menor error aleatorio, y por tanto mayor fiabilidad de la medición realizada. Un valor del $CF \geq 0,7$ será el mínimo requisito exigible a cualquier cuestionario que pretenda evaluar la CVRS de forma fiable. Un CF de 0,7 supone que la puntuación observada coincide en un 70% con la puntuación verdadera de los individuos evaluados, mientras que un 30% de la misma se debe a errores aleatorios de medida. Para determinados usos de los instrumentos de CVRS (p. ej., decisiones en pacientes individuales a partir de sus puntuaciones), el valor mínimo exigible para el CF será por supuesto mayor. Valores entre 0,9 y 0,95 se pueden considerar óptimos, ya que implican únicamente entre un 10% y un 5% de error en la medida²⁴.

Habrà muchas situaciones en las que administrar una segunda vez el cuestionario de CVRS no será posible o recomendable. En aquellos casos en los que, por ejemplo, el constructo evaluado experimente variaciones sustanciales entre el test y el retest, el CF, bajo sin lugar a dudas, no será fruto de una pobre fiabilidad del cuestionario, sino reflejo del cambio producido en el constructo entre ambas administraciones. En tales circunstancias el usuario del cuestionario de CVRS estará interesado en determinar la relación entre las puntuaciones observadas y verdaderas en única administración del instrumento.

Los procedimientos diseñados para estimar la fiabilidad a partir de una única administración del cuestionario se conocen, de forma general, como métodos de análisis de la consistencia interna del ins-

trumento^{9,10}. Destacan dos tipos de estimaciones diferentes: la obtenida mediante la prueba de las dos mitades¹⁰ y el coeficiente alfa de Cronbach^{10,25}.

La prueba de las dos mitades consiste en dividir el cuestionario administrado a los pacientes en dos partes, A y B (p. ej., aleatoriamente, asignando los ítems pares a una mitad y los impares a la otra); seguidamente se puntúan las dos mitades de forma separada para cada paciente y se calcula el coeficiente de correlación de Pearson (r_{AB}) entre ambas series de puntuaciones. El CF para el instrumento se obtiene finalmente mediante la siguiente fórmula: $2r_{AB}/1 + r_{AB}$ ¹⁰.

La mayor limitación de la prueba de las dos mitades es que la utilización de métodos diferentes para dividir el cuestionario en dos partes puede producir también diferentes estimaciones del coeficiente de fiabilidad. El cálculo del coeficiente alfa de Cronbach resuelve estos problemas. El coeficiente alfa²⁵, que puede tomar valores entre 0 y 1, se calcula a partir de la matriz de covariancias de las respuestas a los ítems del cuestionario. Una valor próximo a 1 indica una alta homogeneidad o consistencia en las respuestas proporcionadas a dichos ítems. Una elevada consistencia interna en las respuestas a sus ítems es considerada prueba suficiente de la elevada fiabilidad del instrumento.

Validez

Se acaba de ver que mediante la recogida de datos empíricos es posible determinar hasta qué punto los resultados obtenidos por un cuestionario de medida de la CVRS son estables y reproducibles.

Antes de confiar plenamente en el instrumento y utilizarlo en la práctica se deberá, sin embargo, comprobar una última característica crucial del mismo: su validez²⁰. El concepto de «validez» de

las puntuaciones de un cuestionario se refiere a la utilidad de las inferencias que de estas puntuaciones se pueden extraer⁹. ¿Mide realmente el cuestionario lo que pretende medir? Establecer un proceso de validación del instrumento será necesario para responder a esta pregunta.

El concepto «validación» hace referencia al proceso mediante el cual se recoge evidencia empírica que dé apoyo al uso de las puntuaciones observadas para una propuesta determinada¹⁰. Tradicionalmente se consideran tres aproximaciones diferentes en el proceso de validación: validación de contenido, de criterio y de constructo⁹⁻¹¹.

El establecimiento de la validez del contenido de un cuestionario supone determinar si realmente las preguntas o afirmaciones que éste formula representan la concreción observable del constructo que está siendo evaluado. Típicamente se lleva a cabo mediante la colaboración de expertos en el área que se pretende evaluar. Estos expertos examinan el contenido del instrumento y juzgan hasta qué punto es representativo de aquello que se pretende medir.

La validación de criterio, conocida también como determinación de la validez predictiva, supone un estudio de la relación entre las puntuaciones observadas en el cuestionario y un criterio externo que define adecuadamente y de forma independiente aquello que se pretende medir con el instrumento. Una vez el criterio ha sido determinado, la validez de la predicción se determina a partir del coeficiente de correlación entre las puntuaciones del cuestionario y los valores del criterio. El tamaño de la correlación se emplea como indicador directo de la validez de criterio del instrumento en cuestión. Los mayores problemas de esta aproximación radican en la determinación de un criterio apropiado. Habitualmente es difícil encontrar en el área de CVRS un cri-

terio adecuado para realizar comparaciones con el cuestionario evaluado. Es habitual utilizar esta aproximación para validar aquellos cuestionarios desarrollados con la intención de sustituir o realizar funciones similares a otros cuestionarios ya existentes. El razonamiento es sencillo: si el nuevo cuestionario evalúa «lo mismo» que el cuestionario ya existente (el nuevo cuestionario puede ser más corto, o recoger nuevos aspectos de la CVRS no considerados en el «antiguo»), la asociación entre las puntuaciones obtenidas por ambos instrumentos en los mismos pacientes tiene que ser elevada.

La determinación de la validez de constructo de un cuestionario de CVRS resulta aplicable siempre que el usuario desea extraer inferencias que no pueden ser adecuadamente representadas por un simple criterio o completamente evaluadas a partir de su contenido. La validación de constructo supone realizar una serie de estudios empíricos que pongan a prueba hipótesis específicas sobre cómo los pacientes que difieren en el constructo de interés, lo hacen a su vez en otras variables relacionadas. Si bien los procedimientos para la validación de constructo de un cuestionario son múltiples, a menudo se emplean técnicas de diferenciación de grupos conocidos, el análisis factorial o el análisis de la matriz multirrasgo-multimétodo^{9-11,20}. La lógica subyacente a todos los métodos es, en cualquier caso, muy similar y se basa en el establecimiento hipotético de las relaciones existentes entre el constructo de interés y otras variables relacionadas con el mismo. Por ejemplo, para el caso de la técnica de diferenciación de grupos conocidos, sería más que razonable establecer la siguiente hipótesis: los pacientes con artritis reumatoide presentan más dolor que las personas «sanas» de la población general. La hipótesis establece claramente la

correspondencia entre el constructo dolor y la variable «ser o no ser paciente con artritis reumatoide». Una manera de determinar la validez del cuestionario de dolor utilizado como ejemplo en la tabla 1 consistiría en administrar las preguntas del mismo a 2 grupos de personas: «pacientes con artritis» y «no pacientes»; si el cuestionario mide lo que tiene que medir, dolor, y la hipótesis no es incorrecta (ciertamente improbable en este caso), los pacientes con artritis tienen que mostrar mayores puntuaciones (mayor dolor por tanto) que los no pacientes. De no ser así, de no haber diferencias significativas entre las puntuaciones de ambos grupos, se dirá que el cuestionario no es válido para medir dolor, o que la hipótesis no es correcta. Dado que todas las formas de validación (contenido, criterio y constructo) implican de una u otra manera el establecimiento de hipótesis a priori sobre las relaciones entre las puntuaciones observadas y terceras variables, algunos autores han argumentado que únicamente existe una única forma de validez: la validez de constructo^{9,10}.

En cualquier caso, el proceso de validación, sea de la naturaleza que sea, es continuo y no puntual. La existencia de una sola prueba de evidencia de validez del instrumento no suele considerarse suficiente²⁰. Cada nueva ocasión en la que el instrumento es utilizado supone una nueva oportunidad para poner a prueba las expectativas que sobre él se tienen.

Una última consideración

Con la validación del instrumento se cierra el proceso psicométrico descrito en la figura 2 y el proceso subyacente a la medición de la CVRS a través de cuestionarios. Como se indicaba en la introducción, la comprensión de la metodología implicada en el proceso facilitará, sin duda, el uso y la interpretación correcta de las pun-

tuaciones proporcionadas por dichos cuestionarios.

A pesar de que las consideraciones expuestas son las básicas que cualquier usuario debería tener en cuenta a la hora de juzgar la adecuación de un cuestionario de CVRS en cualquier circunstancia, existen otros aspectos adicionales que también deberían ser considerados al evaluar de forma exhaustiva este tipo de instrumentos²⁰. Cuestiones como la sensibilidad del instrumento para detectar cambios en la CVRS clínicamente (y no sólo estadísticamente) significativos²⁶, o la adecuación de la traducción y/o adaptación del cuestionario en el caso de que originalmente fuera desarrollado en otro idioma y/o cultura²⁷ deberían ser foco de atención, no sólo del investigador experto en la materia, sino también del usuario más modesto simplemente interesado en la interpretación de las puntuaciones observadas obtenidas por sus pacientes en un cuestionario de CVRS.

Bibliografía

1. Leplege A, Hunt S. The problem of quality of life in medicine. *JAMA* 1997; 278: 47-50.
2. Patrick DL, Bergner M. Measurement of health status in the 1990s. *Ann Rev Public Health* 1990; 11: 165-183.
3. McDowell I, Newell C. *Measuring health: a guide to rating scales and questionnaires* (2.^a ed.). Nueva York: Oxford University Press, 1996.
4. Badia X, Prieto L, Roset M, Díez-Pérez A. Desarrollo del cuestionario clínico ecos-16 para la evaluación de la calidad de vida en pacientes con osteoporosis. *Med Clin (Barc)* 2000; 114 (Supl 3): 68-75.
5. McGee H, Hannah M, O'Boyle CA, Hickey A, O'Malley K, Joyce CRB. Assessing the quality of life of the individuals: the SEIQOL with a healthy and gastroenterology unit population. *Psychol Med* 1991; 21: 749-759.
6. Badia X, Salamero M, Alonso J. La medida de la salud. Guía de escalas de medición en español. Barcelona: EDIMAC, 1999.
7. Badia X, Carne X. La evaluación de la calidad de vida en el contexto de los ensayos clínicos. *Med Clin (Barc)* 1998; 110: 550-556.
8. Badia X, Guerra L, Garcia M, Podzamczak D. La evaluación de la calidad de vida en los pacientes con infección por el VIH y sida. *Med Clin (Barc)* 1999; 112: 739-744.
9. Nunnally JC, Bernstein IH. *Psychometric theory* (3.^a ed.). Nueva York: McGraw-Hill, 1994.
10. Crocker L, Algina J. *Introduction to classical and modern test theory*. Fort Worth: Harcourt Brace Jovanovich, 1986.
11. Suen HK. *Principles of test theories*. New Jersey: LEA, 1990.
12. Patrick DL, Erikson P. *Health status and health policy: quality of life in health care evaluation and resource allocation*. Nueva York: Oxford University Press, 1993.
13. Prieto L. La elaboración de un cuestionario. *FMC* 2000; 7: 138-146.
14. Rasch G. *Probabilistic models for some intelligence and attainment tests*. Chicago: Mesa Press, 1993. (Original Edition, The Danish Institute For Educational Research, 1960.)
15. Hambleton RK, Swaminathan H, Rogers HJ. *Fundamentals of item response theory*. Newbury Park: Sage Pub., 1991.
16. Haley SM, McHorney CA, Ware JE Jr. Evaluation of the MOS SF-36 physical functioning scale (PF-10): I. Unidimensionality and reproducibility of the Rasch item scale. *J Clin Epidemiol* 1994; 47: 671-684.
17. Prieto L, Alonso J, Ferrer M, Antó JM. Are results of the SF-36 Health Survey and the Nottingham Health Profile similar?: a comparison in COPD patients. *J Clin Epidemiol* 1997; 50: 463-473.
18. McHorney CA, Haley SM, Ware JE. Evaluation of the MOS SF-36 Physical Functioning Scale (PF-10): II. Comparison of relative precision using Likert and Rasch Scoring Methods. *J Clin Epidemiol* 1997; 50: 451-461.
19. Prieto L, Alonso J, Lamarca R, Wright BD. Rasch measurement for reducing the items of The Nottingham Health Profile. *J Outcomes Measurement* 1998; 2: 285-301.
20. Streiner DL, Norman GR. *Health measurement scales. A practical guide to their development and use*. Oxford: Oxford University Press, 1989.
21. Gorsuch RL. *Factor analysis* (2.^a ed.). Hillsdale: LEA, 1983.

22. Hays R, Stewart A. The structure of self-reported health in chronic disease patients. *Psychol Assess* 1990; 2: 22-30.
23. Juniper EF, Guyatt GH, Streiner DL, King DR. Clinical impact versus factor analysis for quality of life questionnaire construction. *J Clin Epidemiol* 1997; 50: 233-238.
24. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res* 1995; 4: 293-307.
25. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951; 16: 297-334.
26. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol* 2000; 53: 459-468.
27. Herdman M, Fox-Rushby J, Badia X. A model of equivalence in the cultural adaptation of HRQoL instruments: the universalist approach. *Qual Life Res* 1998; 7: 323-335.