

## ¿Cómo validar un instrumento de medida de la salud?

### *How is an instrument for measuring health to be validated?*

A. Carvajal<sup>1</sup>, C. Centeno<sup>2</sup>, R. Watson<sup>3</sup>, M. Martínez<sup>2</sup>, Á. Sanz Rubiales<sup>4</sup>

#### RESUMEN

Cada vez es más necesario disponer de instrumentos de medida en el ámbito de la salud que se puedan utilizar en la práctica clínica e investigación. Para garantizar la calidad de su medición es imprescindible que los instrumentos sean sometidos a un proceso de validación. Este proceso consiste en adaptar culturalmente un instrumento al medio donde se quiere administrar y comprobar sus características psicométricas como: fiabilidad, validez, sensibilidad y factibilidad. Hay instrumentos de medida del ámbito de la salud, disponibles en otros idiomas pero sin validar al español. Además la metodología para validar un instrumento es poco conocida por profesionales sanitarios, lo que explica el uso indiscriminado de instrumentos sólo adaptados o validados de manera poco consistente.

El objetivo de esta revisión es realizar una puesta al día del proceso de validación de un instrumento de medida de la salud y en qué consiste de manera práctica. La accesibilidad de instrumentos de evaluación adaptados culturalmente y validados en distintos idiomas facilitará comparar resultados obtenidos con el mismo instrumento y desarrollar estudios internacionales en distintas culturas.

**Palabras clave.** Validación. Adaptación cultural. Aspectos psicométricos. Fiabilidad. Validez.

#### ABSTRACT

It is increasingly necessary to have a measuring instrument available in the health field that can be used in clinical practice and research. In order to guarantee the quality of their measurements it is essential that the instruments should be subjected to a process of validation. This process consists in adapting the instrument culturally to the setting where its psychometric characteristics are to be administered and checked, such as: reliability, validity, sensitivity and feasibility. There are measuring instruments from the health field available in other languages but that have not been validated into Spanish. Besides, the methodology for validating an instrument is little understood by the health professionals, which explains the indiscriminate use of instruments that have only been adapted or validated in a way that is not very consistent.

The aim of this review is to bring up to date the process of validating an instrument for measuring health, and what it involves, in a practical way. The accessibility of evaluation instruments that have been culturally adapted and validated in different languages will facilitate the comparison of results obtained with the same instrument and the development international studies in different cultures.

**Key words.** Validation. Cultural adaptation. Psychometric aspects. Reliability. Validity.

*An. Sist. Sanit. Navar. 2011; 34 (1): 63-72*

1. Facultad de Enfermería. Universidad de Navarra.
2. Unidad de Medicina Paliativa. Clínica Universidad de Navarra.
3. Escuela de Enfermería. Universidad de Sheffield (Reino Unido).
4. Hospital Universitario del Río Hortega. Valladolid.

Recepción: 17 de septiembre de 2010

Aceptación provisional: 12 de noviembre de 2010

Aceptación definitiva: 15 de diciembre de 2010

#### Correspondencia

Ana Carvajal Valcárcel  
Facultad de Enfermería  
Universidad de Navarra  
Campus Universitario  
31008-Pamplona (Navarra)  
E-mail: acarvajal@unav.es

## INTRODUCCIÓN

En el proceso complejo de valorar el estado de salud o enfermedad el clínico pone a prueba su formación y experiencia. La práctica clínica individual no exige necesariamente el uso de escalas o cuestionarios aunque pueden asistir en el despistaje de procesos o en su seguimiento evolutivo. El cuestionario suprime presuntamente el factor subjetivo del evaluador y ésta es la base de su utilidad en la investigación clínica. La aplicación de una escala, sin embargo, no es un proceso automático ni está exento de dificultades en la práctica. Los instrumentos muchas veces exigen la colaboración del paciente y el que tenga una formación básica o una función cognitiva suficiente para responder a las cuestiones. El modo de preguntar y el lenguaje utilizado son fuentes de sesgos, pero no lo son menos los factores culturales que llevan a que una misma cuestión pueda ser o no válida en un idioma o en otro, o incluso en países diferentes que comparten la misma lengua.

Para utilizar con garantía un cuestionario es necesario comprobar que mide apropiadamente y de modo constante en las mismas circunstancias, que es sensible a los cambios de la situación clínica, que en la práctica no presenta dificultades de aplicación, etc. Toda esta comprobación de sus características es lo que se llama proceso de validación. Este proceso se realiza inicialmente en el idioma original y debe reiterarse cuando se altera la estructura o cuando cambia el idioma en el que se aplica. Validar una versión en idioma diferente al original consiste en adaptar culturalmente el cuestionario al medio donde se quiere utilizar y volver a comprobar que conserva las características psicométricas adecuadas para medir aquello para lo que fue diseñado<sup>1</sup>. La calidad de la metodología utilizada es imprescindible para asegurar que los resultados que se obtienen en el estudio se puedan interpretar adecuadamente y utilizar en la práctica clínica<sup>2</sup>.

Es necesario disponer de instrumentos validados en el ámbito de la salud en el idioma original para poder comparar los

resultados obtenidos con otros estudios nacionales e internacionales que hayan utilizado el mismo instrumento<sup>5</sup>. Existe una monografía en castellano que recopila, actualiza y analiza los instrumentos actualmente disponibles para la medición de la salud, la calidad de vida en la práctica y la investigación clínica; proporciona la información para conocer los principios básicos de la medida de la salud y una guía detallada de los cuestionarios desarrollados, traducidos y validados en español para que los médicos y decisores sanitarios puedan optar, en cada momento y ámbito de aplicación sanitaria, por el instrumento a utilizar<sup>3</sup>.

Con motivo de un proyecto que consistía en la validación del sistema de evaluación de síntomas de Edmonton (ESAS)<sup>4</sup> hemos realizado una puesta al día de qué es el proceso de validación de un instrumento en el ámbito de la salud y en qué consiste. Esta revisión podría ser de utilidad a investigadores que quieran conocer como validar un instrumento y se plantean llevarlo a cabo.

Para validar un instrumento de medida hay que comenzar por adaptar culturalmente el cuestionario al medio donde se quiere utilizar y a continuación volver a medir sus características psicométricas. El objetivo de este trabajo es repasar estos pasos y daremos a conocer los distintos aspectos psicométricos que se pueden medir de un cuestionario.

## LA ADAPTACIÓN TRANSCULTURAL DE UN INSTRUMENTO

El desarrollo de estudios internacionales y la diversidad de culturas en algunos países están aumentando progresivamente<sup>2</sup>. Esto requiere tener accesibles instrumentos de evaluación que estén adaptados culturalmente en los distintos países<sup>5</sup>. La Organización Mundial de la Salud (OMS) recomienda desarrollar y utilizar indicadores de la salud estandarizados. En concreto exige a la investigación multicéntrica internacional tener instrumentos validados en el ámbito de la salud para poder realizar estudios comparativos a nivel internacional<sup>6</sup>.

Algunos autores<sup>7</sup> señalan ciertos criterios para asegurar la calidad del instrumento: traducción, adaptación, validación de las propiedades psicométricas y redacción de las normas del instrumento en los distintos idiomas. Sin embargo la metodología para adaptar cuestionarios es poco conocida por profesionales sanitarios en nuestro medio, lo que explica la existencia y uso indiscriminado de instrumentos adaptados de manera incompleta o traducciones simplemente literarias<sup>8</sup>. Uno de los sesgos tradicionalmente encontrados en instrumentos culturalmente adaptados es la ausencia de equivalencia conceptual de los instrumentos en distintas culturas<sup>7</sup>. Por ello al validar un instrumento es necesario empezar con el proceso de traducción.

### El proceso de traducción

El proceso de traducción de un cuestionario es clave para asegurar la validez de un instrumento en distintas culturas. No puede limitarse a una simple traducción del cuestionario sino que debe seguir una metodología que asegure esa equivalencia<sup>2</sup>. Es un proceso riguroso que consiste en la traducción y adaptación cultural de la versión original a la versión adaptada<sup>9</sup>. El objetivo es conseguir que el instrumento sea equivalente a nivel semántico, conceptual, de contenido, técnico y de criterio en distintas culturas<sup>10,11,12</sup>. La equivalencia semántica asegura que el significado de cada ítem sea el mismo en cada cultura después de la traducción. La equivalencia conceptual nos afirma que el instrumento mide el mismo constructo teórico en cada cultura. La equivalencia de contenido certifica que el contenido de cada ítem es relevante en cada cultura. La equivalencia técnica nos dice que el método de recogida de datos es comparable en cada cultura. Por último, la equivalencia de criterio nos cerciora que la interpretación de la medida se mantiene igual cuando se compara con las normas de cada cultura estudiada<sup>13</sup>. La comprobación de la equivalencia técnica y de criterio de una nueva versión puede requerir de un análisis estadístico posterior, proceso que excede al de la traducción inicial.

No existe una guía estándar para realizar el proceso de adaptación transcultural a otro idioma. La mayoría de los estudios realizados de traducción y adaptación de instrumentos utilizan diferentes métodos que incluyen una o más de las siguientes técnicas: traducción del instrumento, retrotraducción, técnica bilingüe, comité evaluador, y estudio piloto. Estas técnicas son recomendables, sin embargo, ninguna de ellas es completa por lo que se recomienda realizar más de una al mismo tiempo<sup>14,15</sup>.

Los métodos más utilizados son: la traducción lineal, la traducción lineal con estudio piloto y la traducción-retrotraducción con personas monolingües o bilingües. El método que se considera más completo y garantiza una mayor calidad en el proceso de traducción es la traducción-retrotraducción por personas bilingües o monolingües<sup>2</sup>. Este método debe evaluar en primer lugar si el fenómeno que mide el cuestionario existe en la cultura a la que se adapta. A continuación se realizan al menos dos traducciones de la versión original a la lengua de la población diana por traductores bilingües cuya lengua materna sea la población diana. La traducción debe ser conceptual por lo que la persona que traduce debe conocer además del idioma los contenidos y el propósito del cuestionario. A partir de estas traducciones varios evaluadores miden la equivalencia semántica de las distintas versiones y un comité de revisión diseña una primera versión del cuestionario que puede testarse o no en un estudio piloto con pacientes con las mismas características que los pacientes a los que va dirigido el cuestionario. El comité de revisión, en su caso, analiza los resultados obtenidos y diseña una segunda versión del cuestionario. Esta versión se retrotraduce a la lengua original por dos personas bilingües. Estas versiones se comparan con la versión original en equivalencia semántica y el Comité de Revisión decide si se requiere modificar la versión obtenida antes de medir las propiedades psicométricas. El consensuar la última versión con alguno de los autores de la versión original del cuestionario es garantía de calidad del proceso de adaptación del cuestionario<sup>8</sup>.

Una vez finalizado el proceso de traducción se lleva a cabo la validación del instrumento comprobando sus propiedades psicométricas.

## VALIDACIÓN DE CUESTIONARIOS Y ESCALAS

### Características psicométricas de un instrumento

Evaluar las propiedades psicométricas de un instrumento es un criterio esencial para determinar la calidad de su medición<sup>16</sup>. Las dos características métricas esenciales para valorar la precisión de un instrumento son la fiabilidad y la validez<sup>16</sup>. La fiabilidad hace referencia al hecho de medir una variable de manera constante y la validez a que el instrumento mide lo que quiere medir. No todo instrumento que sea fiable es válido. Un instrumento puede ser fiable porque mide una variable de manera constante, pero inválido si no mide el fenómeno que quiere medir<sup>17</sup>. La sensibilidad y factibilidad son otras características métricas que también miden la validez de un instrumento. En la tabla 1 se presentan las diferentes propiedades psicométricas que se pueden medir de un instrumento.

La validación de un instrumento es un proceso continuo y dinámico que va adquiriendo más consistencia cuantas más propiedades psicométricas se hayan medido en distintas culturas, con distintas poblaciones y sujetos<sup>16</sup>.

Actualmente no existe una guía estándar para validar medidas de salud, sin embargo, se utilizan criterios diseñados en ciencias de psicología y educación<sup>18</sup>. La manera de validar un instrumento varía según el tipo y objetivo del instrumento<sup>19,20</sup>. Los instrumentos cuyo objetivo sea recoger información fáctica, relacionada con las acciones que llevan a cabo los sujetos, requerirán que se verifique la validez del contenido por expertos, en cambio, los instrumentos de medida cuantitativos, que valoran la importancia de una variable, requerirán verificar la validez de contenido analizando el concepto expresado en la variable considerada<sup>17</sup>.

### Concepto de fiabilidad

Fiabilidad es la propiedad que designa la constancia y precisión de los resultados que obtiene un instrumento al aplicarlo en distintas ocasiones<sup>17,21</sup>. La fiabilidad evalúa el grado de consistencia en que un instrumento mide lo que debe medir<sup>17,21</sup>. Un instrumento es fiable cuando los resultados son comparables en situaciones similares. La fiabilidad se mide en grados y se expresa en forma de coeficiente de correlación que varía de 0, significando ausencia de correlación, a 1 que es una correlación perfecta<sup>17</sup>. Ningún instrumento presenta el máximo grado de correlación en las distintas situaciones, por ello es necesario determinar el grado de correlación aceptable. Según algunos autores, el margen aceptable para los coeficientes de fiabilidad se sitúa entre 0,7 y 0,9<sup>20,21</sup>. Un coeficiente de fiabilidad alto es claramente deseable cuando las diferencias entre los sujetos son legítimas y esperadas; y esto es lo que suele suceder en los test psicológicos.

No hay que olvidar que una fiabilidad alta no es sinónimo sin más de calidad porque puede faltar lo que es más importante, la validez.

La fiabilidad se puede estimar por cuatro medios: la consistencia interna, la estabilidad, la equivalencia y la armonía interjueces. El método utilizado depende de la naturaleza del instrumento<sup>17,21</sup>.

La consistencia interna mide la homogeneidad de los enunciados de un instrumento indicando la relación entre ellos<sup>17,21</sup>. Este método es el más utilizado para medir la fiabilidad de los instrumentos<sup>21</sup>. Hay varias técnicas para medirla como son la fiabilidad mitad y mitad, el alfa de Cronbach y la técnica de Kuder-Richardson<sup>5</sup>. El más utilizado es el alfa de Cronbach que mide la correlación de los ítems dentro del cuestionario valorando cómo los diferentes ítems del instrumento miden las mismas características<sup>20</sup>. Se considera el indicador más idóneo porque da un único valor de consistencia y proporciona los datos de la técnica de la fiabilidad mitad y mitad. El rango del alfa de Cronbach oscila de 0 a 1. Valores

Tabla 1. Características psicométricas de un instrumento

| ASPECTOS PSICOMÉTRICOS       |  | DESCRIPCIÓN  |
|------------------------------|--|--|
| <b>FIABILIDAD</b>            | <b>Consistencia interna</b>  | <i>Alfa de Cronbach</i><br>Mide el grado de correlación interna entre los ítems  |
|                              |  | <i>Mitad y mitad</i><br>Compara las correlaciones entre las dos mitades de un instrumento                                      |
|                              |  | <i>Kuder-Richarson:</i><br>Obtiene el grado de correlación entre variables dicotómicas   |
|                              |  | <i>Correlación internunciados</i><br>Obtiene el grado de correlación entre cada variable y la puntuación total                 |
| <b>Estabilidad</b>           | <i>Test-retest</i><br>Mide la constancia de las respuestas obtenidas en repetidas ocasiones con los mismos sujetos   |  |
| <b>Equivalencia</b>          |  | Determina la consistencia de las puntuaciones de los instrumentos  |
| <b>Armonía interjueces</b>   |  | Mide el grado de concordancia entre observadores que miden el mismo fenómeno   |
| <b>VALIDEZ</b>               | <b>Validez aparente</b>  | Determina el grado en el que los ítems parecen medir lo que se proponen  |
|                              | <b>Validez de contenido</b>  | <i>Método Delphi</i><br>Método para obtener la opinión de un panel de expertos   |
|                              |  | Modelo de estimación de magnitud<br>Determina la intensidad percibida de un estímulo físico o social                           |
|                              |  | <i>Modelo de Fehring</i><br>Explora si el instrumento mide el concepto que quiere medir con la opinión de un grupo de expertos |
|                              |  | <i>Metodología Q</i><br>Mide la validez de contenido con un grupo de expertos  |
| <b>Validez de criterio</b>   |  | Mide el grado de correlación entre un instrumento y otra magnitud que mida el mismo criterio                                   |
|                              | <i>V.concomitante</i><br>Mide el grado de correlación entre dos medidas del mismo concepto al mismo tiempo en los mismos sujetos   |  |
|                              | <i>V. predictiva</i><br>Mide el grado de correlación entre la medida de un concepto y una medida posterior del mismo concepto. Mide como un instrumento predice una evaluación |  |
| <b>Validez de constructo</b> | <i>Convergente-divergente</i><br>Mide si el instrumento correlaciona con variables esperables y no correlaciona con las que no se esperan                                      |  |
|                              | <i>Análisis factorial</i><br>Reduce un número de variables a factores de variables para distinguir las dimensiones subyacentes que establecen las relaciones entre los ítems   |  |
|                              | <i>Validez discriminante</i><br>Mide el grado del instrumento para distinguir entre individuos que se espera que sean diferentes   |  |
| <b>SENSIBILIDAD</b>          |  | Habilidad del instrumento para reflejar cambios en el estado de salud debido a una intervención conocida                       |
| <b>FACTIBILIDAD</b>          | <i>Porcentaje de respuestas</i><br><i>Tiempo de cumplimentación</i><br><i>Percepción de los pacientes y los profesionales</i>  | Mide si el cuestionario es asequible para utilizarlo en el campo que se quiere utilizar  |

elevados denotan una mayor consistencia interna<sup>21</sup>. Según George y Mallery<sup>22</sup>, el alfa de Cronbach por debajo de 0,5 muestra un nivel de fiabilidad no aceptable; si tomara un valor entre 0,5 y 0,6 se podría considerar como un nivel pobre; si se situara entre 0,6 y 0,7 se estaría ante un nivel débil; entre 0,7 y 0,8 haría referencia a un nivel aceptable; en el intervalo 0,8-0,9 se podría calificar como de un nivel bueno, y si tomara un valor superior a 0,9 sería excelente.

La estabilidad mide la constancia de las respuestas obtenidas en repetidas ocasiones, en las mismas condiciones y con los mismos sujetos. La estimación de la estabilidad se puede obtener con la técnica del test-retest<sup>21</sup>. El método es sencillo y fácil de realizar; consiste en utilizar el instrumento en una misma muestra de sujetos en distintas ocasiones y comparar los resultados de ambos momentos. La determinación de los intervalos de tiempo entre las mediciones depende de los tipos de variables. Algunos autores sugieren un intervalo de dos a cuatro semanas en aquellos instrumentos que midan variables estables<sup>17</sup>. Sin embargo, en los instrumentos que miden variables cambiantes, como son aspectos físicos y psicológicos, es más difícil de determinar el tiempo, porque los resultados pueden ser distintos aunque el instrumento sea estable<sup>21</sup>. Para ello, se determina el intervalo de tiempo en el que se espera que no haya cambios en los sujetos que se evalúan. Otra limitación en esta técnica es que los sujetos en la segunda administración del cuestionario puedan recordar las respuestas de la primera cumplimentación del mismo. Esto puede llevar a un coeficiente de correlación erróneo elevado. Por último, hay que tener en cuenta que los sujetos en la segunda administración puedan cumplimentarlo con menos atención que en la primera o no aceptan que se les administre el cuestionario en dos ocasiones. Estas limitaciones se tienen que tener en cuenta al utilizar esta técnica, sin embargo, no son un impedimento para aplicarla en aquellos instrumentos que se considere necesario medir la estabilidad<sup>21</sup>. Para valorar la estabilidad mediante test-retest se puede usar el coeficiente de correlación de Pearson, de

Spearman o el coeficiente de correlación intraclase (CCI). El coeficiente de Pearson, método estadístico paramétrico, se suele utilizar para medir la correlación entre dos variables cuantitativas, sin embargo, cuando las variables son nominales u ordinales está más indicada la correlación de Spearman-Brown<sup>17</sup>. Estas correlaciones están basadas en el análisis de regresión lineal entre los ítems. El coeficiente de correlación intraclase (CCI), además, detecta el cambio de las puntuaciones y permite obtener un solo coeficiente a pesar del número de aplicaciones del cuestionario<sup>5</sup>.

La equivalencia es otra manera de medir la fiabilidad de un instrumento cuando se dispone de dos o más versiones del mismo test. Se mide el grado de correlación entre las versiones aplicándolas sucesivamente a los sujetos en un mismo tiempo. El resultado se obtiene correlacionando las puntuaciones de las dos formas paralelas del instrumento<sup>17</sup>. Valores de correlación por encima de 0,8 reflejan que los instrumentos son equivalentes<sup>23</sup>.

La armonía interjueces o armonización de las medidas de los diferentes observadores mide el grado de concordancia entre los resultados de dos o más observadores al medir las mismas variables o acontecimientos. Se utiliza cuando se quiere determinar la equivalencia de puntuaciones de diferentes sujetos al cumplimentar el mismo instrumento. Se puede realizar también con el mismo observador en dos ocasiones distintas<sup>17</sup>. Habitualmente se obtiene calculando el coeficiente de correlación de Pearson o Spearman. Hay otras técnicas, para obtener la armonía interjueces, como son el coeficiente de correlación Kappa, análisis de varianza, correlación intraclase, que podrían obtener resultados más fiables<sup>21,24</sup>. Estos coeficientes deben alcanzar puntuaciones por encima de 0,5 y se recomienda llegar a 0,7 para que sean consistentes<sup>25,26</sup>.

### Concepto de validez

La validez explora en qué grado un instrumento mide lo que debería medir, es decir aquello para lo que ha sido diseñada

do<sup>17,21</sup>. La validez es una pieza clave tanto en el diseño de un cuestionario como en la comprobación de la utilidad de la medida realizada<sup>16</sup>. Puede estimarse de diferentes maneras cómo son la validez de contenido, validez de criterios y validez de constructo<sup>21</sup>. Cada una de ellas proporciona evidencias a la validación global del instrumento.

La validez aparente es una forma de validez de contenido que mide el grado en el que los ítems parecen medir lo que se proponen. Es el método más sencillo para medir la validez del instrumento. Se considera la parte del "sentido común" de la validez de contenido que asegura que los ítems del instrumento sean adecuados<sup>27</sup>. Consiste en que algunas personas, expertas o no, digan si consideran relevantes los ítems incluidos en un instrumento. Cuantas más personas intervengan, mayor validez tendrá esta técnica. Medir la validez aparente es importante porque la aceptación de una escala por varias personas da consistencia a la hora de utilizarla. La validez de contenido es un método relevante sobre todo cuando se diseña un instrumento. No es tan importante cuando el instrumento ya ha sido validado anteriormente y utilizado en distintos ámbitos<sup>28</sup>. Sin embargo, cuando se traduce un instrumento a otro idioma, si se supone que los conceptos explorados cambian significativamente de una cultura a otra, puede ser conveniente volver a comprobar la validez aparente.

La validez de contenido evalúa de manera cualitativa si el cuestionario abarca todas las dimensiones del fenómeno que se quiere medir, ya que se considera que un instrumento es válido en su contenido si contempla todos los aspectos relacionados con el concepto que mide<sup>1</sup>. Para ello es necesario tener una idea clara de los aspectos conceptuales que se van a medir<sup>17</sup>. Uno de los problemas que aparece al utilizar esta técnica es que no suele estar disponible un listado del contenido correcto del fenómeno que se va a medir y por lo tanto hay que establecerlo. Esto supone el riesgo de que no se incluyan todas las dimensiones<sup>27</sup>. Los métodos utilizados para medir esta validez son: el método Delphi, el modelo de esti-

mación de magnitud, el modelo Fehring y la metodología Q. En la tabla 1 se explica cada una de ellas. Estas técnicas se pueden utilizar individualmente o combinadas entre sí<sup>23</sup>. En ocasiones es suficiente utilizar una de ellas.

Validez de criterio es el grado de correlación entre un instrumento y otra medida de la variable a estudio que sirva de criterio o de referencia<sup>17,21</sup>. Cuando hay una medida de criterio aceptada por investigadores del campo se le llama estándar o regla de oro y los nuevos instrumentos que miden el mismo concepto se comparan con esa medida<sup>29</sup>. Sin embargo, cuando no existen medidas previas que puedan considerarse como medida estándar, la validez de criterio se mide buscando otro instrumento que sirva de medida comparable<sup>21</sup>. Este instrumento debe estar validado en el idioma que se está validando el nuevo instrumento. Hay dos formas de validez de criterio: la validez concomitante y la validez predictiva. La validez concomitante mide el grado de correlación entre dos medidas del mismo concepto al mismo tiempo y en los mismos sujetos<sup>21</sup>. Puede medir la correlación con el instrumento en global o de cada ítem. Esta técnica se utiliza también para seleccionar los mejores ítems de un instrumento y realizar las modificaciones en el instrumento que se está validando<sup>5</sup>. Es importante estar seguro de que el instrumento utilizado como estándar ha sido validado y no solo reconocido como instrumento estándar por la frecuencia de su uso. En el caso de que no esté validado y exista otro instrumento que sí lo esté se utiliza como validez de criterio el que está validado<sup>21</sup>. Esta validez concomitante se expresa por coeficientes de correlación. La validez predictiva mide el grado de correlación entre un instrumento y una medida posterior del mismo concepto o de otro que está estrechamente relacionado. Mide de qué manera un instrumento predice una evolución o un estado posterior<sup>17</sup>. Por ejemplo, un instrumento que mide aptitudes en estudiantes que comienzan unos estudios puede servir para predecir un grado de éxito en el futuro<sup>21</sup>.

La validez de constructo se entiende como el grado en el que un instrumento mide la dimensión evaluativa bipolar para la que fue diseñada. En este sentido se relaciona con la coincidencia de medida con otros instrumentos que midan la misma dimensión. Esta validez determina la relación del instrumento con la teoría y la conceptualización teórica<sup>17,21</sup>. El proceso para medir la validez de constructo empieza definiendo una dimensión o tema, indicando la estructura interna de sus componentes y su relación teórica con otros instrumentos que midan la misma dimensión. Esto puede ser expresado como hipótesis indicando, por ejemplo, qué correlaciones deben tener con otros instrumentos, qué sujetos deben puntuar más alto o más bajo y qué otros resultados se pueden predecir de las puntuaciones<sup>5</sup>. La validez de constructo es un proceso continuo que contribuye a entender mejor el constructo del instrumento y a realizar nuevas predicciones<sup>5</sup>. Esta validez de constructo se explora especialmente cuando el diseño del instrumento se ha basado en una teoría. En otros procedimientos, la validez de constructo se puede verificar mediante estudios de validez convergente-divergente, de análisis factorial y de validez discriminante<sup>17</sup>. La validez convergente se apoya en la hipótesis de que el instrumento que se valida correlaciona con otra escala de medida y examina si el instrumento se correlaciona con variables de otros instrumentos que deberían estar relacionadas con él. Las variables que se espera que no tengan relación se obtiene con la validez divergente, que se apoya en la hipótesis de que el instrumento no correlaciona con otras escalas. Se podría decir que la validez convergente evalúa la sensibilidad y la divergente la especificidad del instrumento<sup>5</sup>. Se miden con el método multirasgo-multimétodo, de Campbell y Fiske<sup>30</sup>, que emplea conceptos de validez convergente y discriminante.

El análisis factorial es un análisis estadístico multivariante que reduce un conjunto de variables interrelacionadas en un número de variables comunes llamados factores<sup>31</sup>. Por ejemplo, el análisis factorial de un instrumento que evalúa la calidad de

vida puede reagrupar variables que evalúan aspectos físicos, psicológicos, sociales y espirituales. Esta técnica se utiliza para distinguir las dimensiones subyacentes que establecen las relaciones entre los ítems del instrumento<sup>32</sup>. Hay dos tipos de análisis factorial: el exploratorio y el confirmatorio. El exploratorio se utiliza cuando no se conocen previamente los factores definitorios de las variables y el confirmatorio cuando se parte de factores definidos "a priori" y se comprueba la adecuación de los mismos al constructo teórico<sup>5</sup>. La interpretación de un factor se realiza examinando las cargas factoriales de los ítems que incluye. Representan la variabilidad de cada ítem explicada por cada factor<sup>28</sup>. Algunos autores recomiendan tener en cuenta las cargas cuyos ítems muestren al menos un 15% de la varianza común con el factor<sup>33</sup>. Para realizar esta técnica es necesario que exista correlación entre las variables del instrumento. Esto se suele realizar con el test de Bartlett, que indica que existe correlación entre las variables con lo que el análisis factorial tendría sentido; se considera adecuado con un nivel de significación menor de 0,05. El índice de Kaiser Meyer Olkim indica el grado de intercorrelación de las variables; si es mayor de 0,7 se considera factible<sup>34</sup>.

La validez discriminante mide el grado de capacidad del cuestionario para distinguir entre individuos o poblaciones que se espera que sean diferentes<sup>5</sup>. Por ejemplo, se puede esperar que los pacientes con dolor intenso tengan una calidad de vida peor que aquellos con dolor leve. Se puede estimar a través de diferentes métodos estadísticos como: multirazo-multimétodo multivariante, test de Mann-Whitney, coeficiente de correlación<sup>5</sup>.

## Concepto de sensibilidad

La sensibilidad de un instrumento nos muestra la capacidad que tiene de detectar cambios en los atributos o sujetos evaluados después de una intervención<sup>1</sup>. Está relacionada con la magnitud de la diferencia en las puntuaciones del sujeto que ha mejo-



rado o empeorado y las que no han cambiado su situación<sup>32</sup>. Se mide con el estadístico “tamaño del efecto” para evaluar la sensibilidad al cambio relacionando la media de las diferencias entre las puntuaciones antes y después de la intervención con la desviación estándar de la puntuación antes de la intervención. Es importante conocer en los instrumentos relacionados con la salud la diferencia mínima relevante para conocer la eficacia de los tratamientos<sup>29</sup>. Esta característica también es importante para estimar el tamaño muestral de un estudio, cuanto mayor sea la sensibilidad del instrumento menor será el tamaño de la muestra necesario<sup>1</sup>.

### Concepto de factibilidad

La factibilidad mide si el cuestionario es asequible para utilizarlo en el campo que se quiere utilizar. Los aspectos que habitualmente se evalúan son: el tiempo que se requiere para cumplimentarlo, la sencillez y la amabilidad del formato, la brevedad y claridad de las preguntas así como el registro, la codificación y la interpretación de los resultados<sup>1,5</sup>. Esta característica es preciso medirla en distintas poblaciones para conocer si el instrumento es adecuado en sujetos en diferentes situaciones<sup>5</sup>. Se obtiene mediante el cálculo del porcentaje de respuestas no contestadas, y el tiempo requerido para rellenarlo. La factibilidad también se puede medir evaluando la percepción del paciente respecto a la facilidad de usar el cuestionario, así como la percepción del profesional respecto a su utilidad en la práctica clínica.

### CONSIDERACIONES FINALES

En este trabajo hemos realizado una revisión del proceso de validación de instrumentos de medida de salud. Para llevar a cabo este proceso se requiere conocimiento de los distintos aspectos psicométricos, rigor y ayuda estadística para explotar los datos de la mejor manera posible. Para asegurar que los instrumentos utilizados en estudios de investigación

midan lo que queremos medir en distintas poblaciones y sean aplicables en la práctica habitual de nuestro trabajo, es necesario que sean sometidos a un proceso de adaptación y validación de las propiedades psicométricas. Evaluar estas propiedades es criterio esencial para determinar la calidad de su medición. No podemos olvidar que la validación de un instrumento no se obtiene de manera global sino por grados en distintos aspectos del instrumento a través de distintos métodos. Además un instrumento adquiere consistencia en su validez cada vez que se utiliza<sup>21</sup>. Aunque un instrumento se haya validado en una población, es importante medir sus propiedades psicométricas cuando se utiliza en otras áreas o poblaciones<sup>16,21</sup>. Se puede decir que es un proceso continuo que nunca se acaba.

### BIBLIOGRAFÍA

1. ARGIMON JM, JIMÉNEZ J. Validación de cuestionarios. Argimon J, Jiménez J. Métodos de investigación clínica y epidemiológica. Madrid: Elsevier 2004.
2. MANEESRIWONGUL W, DIXON J. Instrument translation process: a methods review. *J Adv Nurs* 2004; 48: 175-186.
3. BADÍA X, ALONSO J. Ed. La medida de la salud: guía de escalas de medición en español. 4ª edición. Barcelona: Editec 2007.
4. BRUERA E, KUEHN N, MILLER M, SELMSER P, MACMILLAN K. The Edmonton Symptom Assessment System (ESAS): A simple method for the assessment of palliative care patients. *J Palliative Care* 1991; 7: 6-9.
5. McDOWELL I, NEWELL C. Measuring health: a guide to rating scales and questionnaires. New York: Oxford University Press 1996.
6. PATRICK D, ERICKSON P. Health Status and Health Policy. Quality of life in health care: evaluation and resource allocation. New York: Oxford University Press 1993.
7. HUI C, TRIANDIS H. Measurement in cross-cultural psychology: A review and comparison of strategies. *J Cross-Cult Psychol* 1985; 16: 131-152.
8. AGRA Y. Adaptación transcultural y validación al español del Rotterdam Symptom Checklist para medir calidad. Universidad Autónoma de Madrid, Madrid: 1997.

9. CHWALOW A. Cross-cultural validation of existing quality of life scales. *Patient Educ Couns* 1995; 26: 313-318.
10. CELLA D, HERNÁNDEZ L, BONOMI A, CORONA M, VAQUERO M, SHIOMOTO G et al. Spanish language translation and initial validation of the functional assessment of cancer therapy quality-of-life instrument. *Med Care* 1998; 36: 1407-1418.
11. DUNCKLEY M, HUGHES R, ADDINGTON-HALL J, HIGGINSON I. Translating clinical tools in nursing practice. *J Adv Nurs* 2003; 44: 420-426.
12. HILTON A, SKRUTKOWSKI M. Translating instruments into other languages: development and testing processes. *Cancer Nurs* 2002; 25:1-7.
13. FLAHERTY J, PATHAK D, MITCHELL T, WINTROB R, RICHMAN J, BIRZ S. Developing instruments for cross-cultural psychiatric research. *J Nerv Ment Dis* 1988; 176: 257-263.
14. BRISLIN R. Wording and translation of research instruments. En: Lonner W, Berry J (Eds). *Field Methods in Cross-Cultura*. London: Research Sage 1986.
15. DAUPHINEE S et al. Translating health status questionnaires and evaluating their quality: the IQOLA project approach. *J Clin Epidemiol* 1998; 913-923.
16. GÓMEZ BENITO J, HIDALGO M. La validez en los tests, escalas y cuestionarios. La sociología en los escenarios 8 (revista electrónica). Centro de Estudios de Opinión 2002. (Consulta febrero 2011). <http://aprendeonline.udea.edu.co/revistas/index.php/ceo/article/viewFile/1750/1370>
17. FORTIN M, NADEAU M. La medida de investigación. Fortin MF (Ed). *El proceso de investigación de la concepción a la realización*. México: McGraw-Hill Interamericana 1999.
18. STEWART A, HAYS R, WARE J. Methods of validity MOS Health Measures. Stewart AL and Ware JE (Eds). *Measuring Functioning and Well-Being*. North Carolina Duke University Press 1993.
19. ANASTASI, A. Envolving concepts for test validation. *Annu Rev Psychol* 1986; 37: 1-15.
20. NUNNALLY J. *Psychometric theory*. New York: McGraw-Hill 1978.
21. POLIT D, HUNGLER B. *Nursing research: principles and methods*. Philadelphia. JB Lippincott & Co 1999.
22. GEORGE D, MALLERY, P. *SPSS/PC+ step by step: A simple guide and reference*. Belmont, CA. Estados Unidos Wadsworth Publishing Company 1995.
23. GUIRAO-GORIS J. *Investigación en nomenclatura enfermera*. Guirao-Goris J, Camaño R, Cuesta A (Eds). *Diagnóstico enfermero. Categorías, formulación e investigación*. Sevilla Ediversitas Multimedia 2001.
24. LANDIS J, KOCH G. The measurement of observer agreement for categorical data. *Bioestatistics Technical Report*. Department of Bioestatistics. Michigan: University of Michigan 1976.
25. STEINER D, NORMAN G. Basic concepts, in health measurement scales. Steiner D, Norman G. (Eds). *A practical guide to their development and use*. New York: Oxford University Press 2003a.
26. STEINER D, NORMAN G. Reliability, in health measurement scales. Steiner D, Norman G (Eds). *A practical guide to their development and use*. New York: Oxford University Press 2003b.
27. GOULD J. Medicine's core values. Profession should not have to make decisions concerning rationing. *Br Med J* 1994; 17:1657.
28. AGRA Y, BADÍA X. Spanish version of the Rotterdam Symptom Check List: Cross-cultural adaptation and preliminary validity in a sample of terminal cancer patients. *Psycho-Oncol* 1998; 7: 229-239.
29. BADIA X, SALAMERO M, ALONSO J. *La medida de la salud: Guía de escalas de medición en español*. Barcelona Edimac 1999.
30. CAMPBELL D, FISKE D. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol Bull* 1959; 56: 81-105.
31. WATSON R, THOMPSON D. Use of factor analysis. *J Adv Nurs literature review*. *J Adv Nurs* 2006; 55: 330-341.
32. BADIA X, LIZÁN L. *Estudios de Calidad de Vida*. Martín Zurro, Cano Pérez J (Eds). *Atención Primaria. Conceptos, organización y práctica clínica*. Elsevier 2003.
33. STEVENS J. *Applied multivariate statistics for the social sciences*. New Jersey Lawrence Erlbaum 1992.
34. BISQUERRA R. *Métodos de Investigación Educativa*. Guía practica. CEAC Barcelona 1989.