

Medición en salud: Algunas consideraciones metodológicas

ANA M. ALARCÓN M^a, Sergio Muñoz N^b.

Some methodological issues about measurements in health

Measurement process is a part of everyday work of health professionals. Technological advances have created precise and sensitive instruments whose results are highly accurate and cast little doubts. However, health care has diversified, becoming a multidimensional and complex task. This means that new aspects, that are difficult to measure, must be addressed, such as health perception by patients or cost usefulness ratios of health actions. Therefore, clinicians and researchers must have a good perception of the quality of measures that they use in their daily work, to make the correct diagnostic, therapeutic or prognostic decisions. This article explores the operational concepts that help in the development and assessment of measuring instruments in terms of validity and truthfulness (Rev Méd Chile 2008; 136: 125-30).

(Key words: *Delivery of health care; Research design; Statistics as topic*)

Recibido el 30 de agosto, 2006. Aceptado el 13 de marzo, 2007.

Departamento de Salud Pública, Facultad de Medicina, Centro de Investigación y Gestión para la Evidencia en Salud (CIGES), Universidad de La Frontera, Temuco, Chile.

^aPhD en Antropología Médica

^bPhD en Bioestadística

El proceso de medición forma parte de la práctica diaria de todos los profesionales de salud; está presente por ejemplo, en la evaluación de la capacidad funcional de los adultos mayores, en la detección de actitudes hacia el beber problema, o en la valoración del dolor subjetivo de un paciente^{1,2}. La aplicación de cuestionarios o escalas generan mediciones que muchas veces determinan importantes decisiones pronósticas y terapéuticas; sin embargo, en la mayoría de los casos se desconoce el origen y capacidad real del instrumento para evaluar la condición bajo estudio.

Medición es simplemente el proceso de asignar valores a ciertos eventos de la realidad³. La dificultad del proceso radica al menos en dos aspectos: que el valor represente realmente el evento que se quiere medir, y que el evento sea expresado en toda su complejidad. En medicina, los instrumentos para determinar la condición biológica de un paciente han experimentado tal desarrollo tecnológico, que han logrado disminuir la incertidumbre de los resultados de una medición⁴. No obstante, es la medición de salud como fenómeno psicosocial que requiere refinar la precisión de sus instrumentos, dado que el objetivo principal en este campo de la salud es recoger, en forma válida y confiable, la percepción subjetiva de los pacientes.

Este artículo expone, desde una perspectiva metodológica, algunos conceptos esenciales y

Correspondencia a: Ana M. Alarcón M. Departamento Salud Pública, Facultad de Medicina-CIGES, Universidad de La Frontera. Montt 112, Temuco. Fono: 45- 325740. Fax: 45-325741. E mail: amalarc@ufro.cl

operativos tanto para la construcción, como para la evaluación de instrumentos de medición en salud, particularmente escalas y cuestionarios utilizados en estudios de población. Se abordarán temas tales como: la generación y selección de ítems que configuran el objeto de medición, ciertas consideraciones respecto del diseño de instrumentos, y finalmente algunas indicaciones para la evaluación de la calidad de un instrumento de medición; este último aspecto es esencial para quienes utilizan cuestionarios o escalas ya diseñadas en otras partes del mundo. Su propósito es contribuir con algunos elementos teóricos básicos para guiar al investigador clínico a tomar decisiones basadas en mediciones válidas y confiables.

Generación y selección de ítems de un instrumento de medición en salud. La etapa inicial en la construcción de un cuestionario o escala que tenga por objetivo recoger la percepción u opinión de los pacientes, implica los siguientes pasos: a) definir con máxima precisión el objeto de la medición, b) atribuir a los objetos o eventos de salud ciertos indicadores que expresen conceptualmente las propiedades del objeto que se pretende medir, y c) reflejar la variabilidad entre los sujetos que están siendo objeto de medición. Habitualmente el investigador parte de ideas que provienen de su experiencia clínica, del análisis de literatura científica, y de diálogos con sus pacientes y colegas, de esta forma produce una conceptualización que le permite identificar dimensiones y variables de su objeto de estudio.

Este proceso implica la construcción y reconstrucción del concepto que se quiere medir en la población; tomemos como ejemplo la definición del concepto "calidad de vida", y la identificación de sus propiedades manifiestas, tales como, "estar bien", "sentirse feliz", "dormir bien", "visitar amigos", entre otras; son justamente estas propiedades o características las que contribuirán a reconstruir el concepto "calidad de vida" como objeto de medición.

En esta fase, que podríamos llamar exploratoria, la incorporación de métodos y técnicas cualitativas de investigación tales como, entrevistas en profundidad, observaciones no estructuradas, y grupos focales constituyen herramientas esenciales para indagar en la condición subjetiva de los pacientes, produciendo descripciones y explica-

ciones a procesos, contribuyendo a delimitar marcos teóricos, y mejorando la comprensión del contexto en que se produce el evento de salud^{5,6}. En consecuencia, son las representaciones de los sujetos, más las propias del investigador clínico, las que se configuran como ítems en un cuestionario. Estas formas de obtención de ítems equivalen a las fuentes de datos necesarios para lograr una muestra representativa de los ítems.

Los desafíos inherentes a esta etapa son, determinar el universo de ítems o preguntas que representen totalmente al evento, establecer sus límites, y discriminar lo que pertenece o no, al evento en estudio. El tema es particularmente complejo porque no existe más herramienta que el criterio, experiencia y sensibilidad teórica del investigador clínico para reunir el máximo de propiedades observables del evento que pretende estudiar.

Consideraciones sobre el diseño del instrumento de medición. El diseño de un instrumento de medición –sea éste escala o cuestionario–, depende de múltiples factores entre los que podemos mencionar: el objetivo para el cual se propone el instrumento, los recursos financieros para su construcción y prueba, acceso a la información, edad y educación de los pacientes quienes serán objeto de medición, y el nivel de precisión de la medida que se pretende obtener. El conjunto de estos factores determinará la extensión del instrumento, el formato, y su presentación; sin embargo lo más relevante en el diseño es la decisión del investigador clínico sobre el nivel de medida de sus variables. Por ejemplo, es diferente decirle a un paciente –cualquiera sea su patología–, que construya una lista con sus síntomas, que elija de una lista aquellos que presenta con mayor frecuencia, o que ordene de una lista que el médico le presenta, aquellos que le provocan mayor molestia. La decisión por una u otra forma dependerá tanto de los intereses del clínico en profundizar sobre los síntomas de sus pacientes, de la aplicación de los resultados; como de la calidad de sus variables, es decir, hasta dónde su variable le permite variar en respuesta. En resumen, lo que se intenta es representar la variabilidad de respuestas entre los pacientes a los que se les aplicará finalmente el instrumento⁷.

En términos operativos significa que para construir un cuestionario, el clínico deberá seguir

los siguientes pasos: definir el concepto y objetivo de la medición, elaborar una lista extensiva de sus ítems, determinar la forma de medición de las respuestas, expresar las preguntas (variables) en palabras apropiadas a la cultura, educación y edad de sus pacientes, construir un borrador de prueba, y finalmente probar el instrumento en una muestra de pacientes. Los resultados de su prueba determinarán: el contenido esencial del instrumento, la dificultad de aplicación y comprensión, y la posibilidad de análisis de las medidas obtenidas.

Calidad y pertinencia de un instrumento: temas de validez y confiabilidad. La determinación de la calidad de un instrumento implica plantearse preguntas tanto de la exactitud con que mide el fenómeno bajo estudio, la estabilidad de la medición o resultado, como de la pertinencia del instrumento para los sujetos o realidad en que se aplican las mediciones. Por ejemplo, un clínico que evalúe la calidad de vida de sus pacientes sea ésta en forma específica o genérica, se enfrenta a la incertidumbre de si los valores obtenidos a través del cuestionario reflejan o no la condición del paciente, si el instrumento es capaz de detectar diferencias entre los sujetos, si cambios en los resultados de la medición en un paciente son consecuencia de cambios en su condición de salud, y finalmente si las preguntas del cuestionario fueron respondidas comprensivamente por el paciente.

Este conjunto de interrogantes aluden directamente a los conceptos de confiabilidad y validez de un instrumento. La estimación de la calidad de un cuestionario o escala es requisito esencial, no sólo en su proceso de construcción, sino también en la utilización de aquellos obtenidos en la literatura médica internacional. Por ejemplo, aunque el investigador clínico haya encontrado algún cuestionario muy apropiado a sus objetivos, validado y utilizado en otras partes del mundo; deberá igualmente validarlo con sus pacientes, ya que su aplicabilidad se limita sólo al contexto social y cultural en que fue creado. Para ello debe utilizar, si corresponde, los mismos procedimientos de determinación de validez y confiabilidad que aquí se exponen, sin embargo, deberá iniciar el proceso de validación del instrumento, con una adaptación idiomática y cultural⁸.

Determinación de la validez de un instrumento.

La validez de un instrumento se refiere a la pertinencia de las inferencias realizadas a partir de las mediciones; en otras palabras, validez significa cuán bien un resultado obtenido refleja al fenómeno bajo estudio, o la capacidad del instrumento para medir la cualidad para la que fue construido^{9,10}. Por ejemplo, un test que mida depresión en un paciente debe dar cuenta de la calidad y magnitud de ese fenómeno, y no del stress de ese paciente; por lo tanto, cuando un investigador clínico construya un cuestionario o intente aplicar alguno ya existente debe –en primer término– examinar con profundidad la naturaleza del evento que desea medir (llámese depresión, stress, angustia, calidad de vida, etc.), y al mismo tiempo, examinar si las variables del cuestionario representan en forma precisa el fenómeno que desea medir.

Existen diversas formas de determinar la validez de un instrumento, no obstante la distinción entre ellas, refleja más una necesidad operativa que conceptual para el investigador. Entre éstas se encuentran la validez de: a) criterio, b) constructo, y c) contenido.

Validez de criterio. Este tipo de validez se asocia generalmente a la capacidad de un instrumento para predecir una conducta o evento de salud. En clínica es común evaluar ciertas conductas de riesgo sanitario y predecir quiénes tienen el riesgo potencial de desencadenar el problema, enfermedades cardiovasculares, depresión entre otros ejemplos. El procedimiento más común para determinar la capacidad predictiva de un instrumento, es demostrar una correlación entre las medidas obtenidas con algún criterio exterior habitualmente llamado “gold standard”, de modo que, si la correlación es alta, la medida obtenida es considerada válida¹¹. Desafortunadamente no siempre se dispone de un criterio de comparación aceptable, por lo tanto el desafío para el clínico será buscar o construir alguno con el que pueda comparar los resultados de su medición, para lo cual deberá conjugar tanto su experiencia clínica como la revisión crítica de literatura.

La selección del criterio se basa en una supuesta relación entre el evento medido y el criterio elegido, como ejemplos clásicos se pueden mencionar los cuestionarios de calidad de

vida específicos, cuyos resultados son contrastados con criterios clínicos elaborados por sociedades profesionales^{12,13}.

Algunas variantes de este tipo de validez se asocian al tiempo en que se realizan las pruebas al paciente. Por ejemplo, si se realizan las mediciones antes que aparezca el evento, y se espera su ocurrencia, estamos frente a una forma de validez predictiva. Estas mediciones son bastantes apropiadas para determinar la asociación entre riesgo sanitario y evento, o predicción de éxito o fracaso de intervenciones.

Por otra parte, si se selecciona un gold standard o criterio, y se compara simultáneamente con los resultados obtenidos a través del test aplicado, estamos frente a una forma de validez concurrente. En este caso se busca saber si la medida obtenida con el test diseñado para tal efecto se acerca o no al criterio, de forma tal que el clínico pueda sustituir el criterio por la medida obtenida a través de su test. No obstante, al aplicar criterios o medidas comparativas, siempre debería preguntarse si éstas son apropiadas a la cultura, situación social o nivel educativo en que se aplican, de modo que constituyan una fuente válida de comparación.

Validez de contenido o aparente. La validez de contenido se refiere al grado en que un instrumento representa el universo o la totalidad de los contenidos del fenómeno que se pretende medir, apunta por tanto, a cuán bien la muestra de preguntas del cuestionario representa el concepto bajo estudio¹⁴. Sin embargo, no se espera que un instrumento cubra detalladamente cada uno de los dominios que conforman un concepto –ello implicaría un instrumento extremadamente largo–, lo que debería esperarse es que el instrumento contenga una muestra representativa de sus dominios y posibles preguntas. Una forma habitual de establecer la validez de contenido es la consulta a expertos en el tema, quienes evalúan a través de una pauta elaborada por el investigador, si los contenidos de un test son pertinentes y adecuados; no obstante, es el investigador clínico, quien finalmente decidirá hasta dónde desea extender el contenido de su cuestionario.

Validez de constructo. El propósito de la validez de constructo es ratificar la teoría que fundamenta la utilización de un test, más que el test en sí mismo;

por ello, la validez de constructo se pregunta cuán bien la medición refleja la naturaleza del concepto o “constructo” que se pretende medir¹⁵. Por ejemplo si el propósito es medir la depresión post-parto (como constructo), el investigador debería preguntarse por las dimensiones teóricas que subyacen en el tema, y las relaciones que se configuran entre éstas. El análisis factorial es el procedimiento estadístico más apropiado para determinar este tipo de validez, ya que puede identificar la cantidad de constructos o factores que existen en el test, y la medida en que cada ítem está relacionado con cada factor¹⁶. En la construcción o evaluación de un instrumento de medición, el análisis de factores se puede utilizar de forma exploratoria o confirmatoria. Usualmente esta técnica estadística es empleada en forma exploratoria, ya que su análisis proporciona al investigador una idea de cuáles y cuántas son las dimensiones que conforman el constructo que está investigando.

Existen dos variantes de validación de constructo denominadas: validez convergente y validez concurrente. En la primera, se correlaciona la medición obtenida a través de un instrumento con otro, que mide el mismo constructo. Tomemos como ejemplo la medición del dolor postoperatorio en un mismo grupo de pacientes; una medida se obtendrá a través de la escala análoga visual, y la otra mediante el test de dolor específico para pacientes quirúrgicos. Si los resultados con una y otra medida, correlacionan positivamente o “convergen”, significa que ambos test miden el mismo constructo.

En la segunda variante –validez concurrente–, los resultados de un test se comparan con los de otro que mide un constructo diferente. Siguiendo el ejemplo anterior, se pueden comparar los resultados de percepción del dolor postoperatorio con capacidad funcional, en el mismo grupo de pacientes. En teoría, ambos test debiesen tener correlaciones negativas o inversas, de lo contrario significa que el test de dolor postoperatorio no es suficientemente específico para medir ese constructo en pacientes quirúrgicos.

CONFIABILIDAD O ESTABILIDAD DE LAS MEDIDAS

La confiabilidad se refiere al grado en que los puntajes observados en una medición están libres de error de medida; se relaciona con la estabilidad

de las observaciones e inferencias realizadas y su precisión¹⁷. Para obtener un alto grado de confiabilidad, el esfuerzo debe dirigirse hacia el control del error en las respuestas de los pacientes, dado que el supuesto es que, si las condiciones son constantes los resultados serán idénticos. Por ejemplo, al aplicar un test de esfuerzo, el clínico debe controlar ciertos factores como: la motivación de sus pacientes, las condiciones del ambiente físico en que aplica la prueba, o la dificultad del test. Algunas formas de estimar la confiabilidad de los datos son: a) análisis de consistencia interna, b) test-retest, y c) confiabilidad interobservador.

Consistencia interna o inter-ítem. Consiste en determinar cuán bien los ítems de un cuestionario o escala correlacionan unos con otros a través de diversos procedimientos estadísticos. Uno de éstos consiste en partir el cuestionario en dos mitades; por ejemplo todas las respuestas pares y todas las impares, y correlacionar ambos grupos de respuestas. Otra forma consiste en correlacionar cada ítem con todos los otros, en este caso se estima la consistencia del ítem en forma individual en relación con todos los demás. El coeficiente alpha de Cronbach es el más apropiado para ítems que tienen tres o más alternativas de respuestas, y Kuder Richardson para ítems que ofrecen respuestas dicotómicas; ambas son pruebas estadísticas que evalúan el ítem y su relación con la extensión del cuestionario¹⁸. El análisis de consistencia interna de un instrumento es frecuentemente realizado con posterioridad al análisis factorial, y contribuye complementariamente a evaluar su validez. Basado en estos resultados el investigador clínico podrá determinar la pertinencia de un ítem en el cuestionario; decisión que va a depender tanto de sus intereses y motivaciones, como de los parámetros estadísticos obtenidos.

Test-retest o medida de estabilidad. Consiste en repetir la prueba dos o más veces a los mismos sujetos con el objeto de correlacionar sus respuestas y determinar su estabilidad. Este método tiene la dificultad que muchas veces las diferencias observadas en las respuestas no se deben sólo al azar, interviniendo por ejemplo, sesgos de memoria de los pacientes, cambios en la forma de plantear las instrucciones, o de las condiciones en que se administra la prueba. Por ejemplo, si se

aplica un cuestionario o escala para evaluar la presencia de stress en ciertos sujetos, probablemente se obtendrán resultados diferentes si el cuestionario se aplica antes o después de las vacaciones de los trabajadores. Es por ello, que el investigador clínico que desee realizar un test-retest debe aplicar su cuestionario resguardando que las condiciones en que se aplica sean similares entre la primera y segunda aplicación.

Confiabilidad interobservador o equivalencia.

Esta forma de estimar confiabilidad se utiliza principalmente en observaciones clínicas diagnósticas, y cuestionarios con preguntas abiertas¹⁹. El objeto es correlacionar los puntajes obtenidos entre distintos observadores, indicando de esta forma, el porcentaje de acuerdo respecto de la medida observada corrigiendo el factor azar. La forma habitual de obtener esta confiabilidad es a través del coeficiente de Kappa; considerando un valor muy bueno a partir de 0,8 y muy malo por debajo de 0,20, así mientras más cerca de cero es el valor, más azar hay presente en las medidas²⁰.

Finalmente es necesario señalar que los elementos discutidos sobre evaluación de calidad de las medidas obtenidas a través de la aplicación de instrumentos, deben cumplir dos requisitos fundamentales; primero, un equipo de trabajo multidisciplinario con el cual discutir contenidos, resultados y tomar decisiones, y segundo la provisión de programas de estadísticos adecuados para realizar las pruebas mencionadas, entre éstos el más utilizado en ciencias de la salud es Stata®, y en ciencias sociales SPSS®.

DISCUSIÓN

El diseño y construcción de un instrumento de medición en salud constituye un proceso complejo en el que confluyen múltiples disciplinas y metodologías, en particular si el propósito es construir escalas de medida. En su fase inicial el aporte de los métodos cualitativos de investigación es fundamental para obtener un mayor acercamiento a los atributos del objeto de estudio; es así como entrevistas y grupos focales con la población objetivo, incrementan el universo de ítems al construir un instrumento de medición²¹. Por otra parte, las herramientas estadísticas evidencian la

validez y consistencia del constructo y sus propiedades a través del análisis tanto de los ítems, como de su interacción en el conjunto del instrumento. Sin embargo, el uso de técnicas cualitativas y estadísticas debe ser apropiado al tipo de medición que el clínico desea realizar, y a la naturaleza del constructo que desea conocer. Un instrumento de medición es válido y confiable cuando es pertinente a la cultura que le dio origen, es amigable y comprensible para quienes lo responden, y de-

muestra cumplir con los estándares estadísticos adecuados al tipo de medición, y a los supuestos que fundamentan su construcción.

Los elementos presentados en este artículo, proveen guías operacionales y analíticas para construir medidas en salud que reflejen con mayor grado de validez y confiabilidad aquellas condiciones y percepciones subjetivas de los pacientes, reconociendo que ninguna medida de este tipo es libre de error.

REFERENCIAS

1. CLANCY C, EISENBERG J. Outcomes research: measuring the end results of health care. *Science* 1998; 282: 246-7.
2. GARRANT A, SCHMITH P, MACKINTOSH A. Quality of life measurement: bibliographic study of patients assessed health outcome measures. *BMJ* 2002; 324: 1417-22.
3. KERLINGER F, LEE H. *Investigación del comportamiento. Métodos de investigación en ciencias sociales*. México: McGraw-Hill, 2002.
4. BANGDIWALA S, MUÑOZ S. Medición de confiabilidad y validez en instrumentos clínicos. *Rev Méd Chile* 1997; 125: 446-73.
5. MILES M, HUBERMAN A. *Qualitative data analysis*. CA: Thousands Oaks Sage Publications, 1994.
6. CRABTREE B, MILLER W. *Doing qualitative research methods for primary care*. CA: Newbury Park Sage Publications, 1992.
7. HANDWERKER WP, BORGATTI S. Reasoning with numbers. En: H. Russell Bernard, ed. *Handbook of Methods in Cultural Anthropology*. CA: Altamira Press, 1998; 593-549.
8. SÁNCHEZ R, ECHEVERRY J. Validación de escalas de medición en salud. *Rev Salud Pública* 2004; 6:302-318. Disponible en: http://www.scielosp.org/scielo.php?pid=S0124-00642004000300006&script=sci_arttext [Consultado el 13 de noviembre de 2006].
9. VIEYTES R. *Metodología de la Investigación en Organizaciones: Epistemología y Técnicas*. Buenos Aires: Editorial de las Ciencias, 2004.
10. STREINER D, NORMAN G. *Health measurement scales a practical guide to their development and use*. NY: Oxford Medical Publications, 2003.
11. NUNNALLY J. *Psychometric theory*. NY: McGraw-Hill, 1978.
12. JUNNIPER E, GUYANTT G, FEENY D. Measuring quality of life in children with asthma. *Qual Life Res* 1996; 5: 35-46.
13. STREINER D. A checklists for evaluating the usefulness of rating scales. *Can J Psychiatry* 1991; 38: 140-8.
14. DOOLEY D. *Social research methods*. NY: Prentice-hall Inc. 1995.
15. CRONBACH L. *Essentials of psychological testing*. NY: Harper and Row, 1984.
16. HARMAN H. *Modern factor analysis*. Chicago: University of Chicago Press, 1976.
17. STREINER DL. Starting at the beginning: An introduction to coefficient alpha and internal consistency. *J Personality Ass* 2003; 80: 99-103.
18. NORMAN G, STREINER D. *PDQ statistics*. Toronto: Decker, 2003.
19. MUÑOZ S, BANGDIWALA S. Interpretation of Kappa and B statistics measures of agreement. *J App Stat* 1997; 24: 105-11.
20. HANLEY B. Standard error of the kappa statistic. *Psychological Bulletin* 1987; 114: 235-41.
21. DE SOUZA M. *El desafío del conocimiento: Investigación cualitativa en salud*. Buenos Aires: Lugar Editorial, 2004.