

VALIDEZ EN LA INVESTIGACIÓN IMAGINOLÓGICA

Drs. Daniela Zavando M^(1,2), Iván Suazo G⁽³⁾, Carlos Manterola D⁽⁴⁾.

1. Universidad Autónoma de Chile, sede Talca.
2. Programa de Doctorado en Ciencias Médicas, Facultad de Medicina, Universidad de La Frontera. Chile.
3. Departamento de Ciencias Básicas Biomédicas. Universidad de Talca. Chile.
4. Departamento de Cirugía y Traumatología, Facultad de Medicina, Universidad de La Frontera. Chile.

VALIDITY IN IMAGING RESEARCH

Abstract: It is common to note that the concept of validity is given added significance in the realm of knowledge and scientific finding dissemination, despite being an attribute of value interpretation of measuring instruments. Improper management of this concept generates confusion and may lead to a false legitimization of scientific inferences. Through this review, the authors intend to contribute to the understanding of this construct analyzing it from a historical and operational perspective in order to allow both the user of the imaging literature and those who report the results of their research to give the concept a semantically correct use as well as to identify some of the practical statistical resources to support this attribute.

Keywords: Diagnosis, Digital radiography, Mammography, Validity.

Resumen: Es habitual observar que al concepto de validez se le otorgue un significado adicional sobre la difusión de conocimientos y conclusiones científicas, no obstante esta sea una propiedad de la interpretación de valores obtenidos de instrumentos de medición. El manejo inadecuado de este concepto genera confusión y puede incurrir en una falsa legitimación de inferencias científicas. Por medio de esta revisión, es intención de los autores colaborar en la comprensión de este constructo desde su perspectiva histórica y operativa, con el objetivo de permitir, tanto al usuario de la literatura imaginológica como a quienes comunican los resultados de sus investigaciones, dar un correcto uso semántico así como distinguir algunos

Financiado parcialmente por proyecto DI09-0060 de la Dirección de Investigación de la Universidad de La Frontera.

Zavando D y cols. Validez en la investigación imaginológica. Rev Chil de Radiol 2010; 16(2); 75-79.

Correspondencia: Dra. Daniela Zavando M. Universidad Autónoma de Chile, sede Talca, Chile. dzavando@yahoo.es

Trabajo recibido el 29 de marzo de 2010, aceptado para publicación el 17 de mayo de 2010.

de los recursos estadísticos prácticos para sostener esta propiedad.

Palabras Clave: Diagnóstico, Mamografía, Radiografía digital, Validez.

Introducción Definiendo validez

Con relativa frecuencia vemos que el concepto de validez es utilizado para significar todos los atributos positivos que posee un instrumento de medición o clasificación, lo cual genera confusión y puede resultar en una legitimación apresurada de un método o instrumento en particular, que bien podría inducir a falsas estimaciones⁽¹⁾. Esta situación no es ajena al quehacer imaginológico y por ello, el foco de interés de esta revisión estará centrado en analizar el concepto desde esta área del conocimiento.

El diagnóstico por métodos imaginológicos involucra un proceso que es dependiente de la interpretación de diferentes elementos que provienen principalmente de:

- El propio método empleado (tales como apreciación de dimensiones lineales/angulares, formas, cambios en la absorción, conocimiento de los fenómenos de proyección, correcta técnica imaginológica)
- La población sobre la que se realizan los análisis
- El *expertise* y conocimientos del especialista.

Como resultado tenemos que el diagnóstico imaginológico se comporta como un constructo complejo. Considerando lo anterior, al momento de abordar el concepto de validez aparece que esta última debe ser analizada desde la perspectiva de la estimación de constructos, tal como se hace desde otras áreas de la ciencia.

La forma más clara de definir validez es "la capacidad de medición o clasificación de un método o instrumento para aquello que fue propuesto, o sea que mida o clasifique lo que efectivamente analizamos y no otra cosa". Al respecto, la imaginología tiene como responsabilidad importante la validación de técnicas

y mediciones en contextos poblacionales diferentes. A modo de ejemplo, en nuestro país es frecuente que la estimación de la maduración ósea sea realizada en radiografía de mano y muñeca, según el método de Greulich y Pyle, cuyos indicadores fueron originalmente reportados a partir de evaluaciones realizadas en un grupo de niños estadounidenses en 1959⁽²⁾, lo que en la actualidad ya pudiera ser ajeno a la propia población norteamericana, considerando los cambios de hábitos nutricionales y el incremento de la obesidad desde el pasado hasta ahora; con mayor razón, podemos esperar diferencias no sólo atribuibles a la temporalidad sino también a las características genéticas de nuestra población. Por supuesto, entre los investigadores del área esperamos estudios de validación con muestras representativas y tamaños muestrales adecuados. En general, podemos comprender la carencia de estudios de validación desde el contexto del desconocimiento de la real significación de validez.

Dado que la psicología es la ciencia que se ha avanzado en esta materia, veremos brevemente la evolución del concepto de validez manifestada por los acuerdos de la American Psychological Association (APA), desde el año 1954 hasta 1999⁽³⁾:

- Se observa una primera fase operacional, donde se impone la máxima que "un instrumento es válido para aquello con que se correlaciona". Esta mirada diferencia 4 tipos de validez: de contenido, predictiva, concurrente y de constructo (APA, 1954), que posteriormente se reducen a 3 (APA, 1974), quedando así la validez concurrente y predictiva reunidas en la validez de criterio; esta visión tripartita será mantenida hasta 1985⁽⁴⁾.
- En la segunda etapa, más teórica, en un intento por lograr una visión más integradora, la validez de constructo se vuelve preponderante debido a que considera también la interacción con otras variables y aumentan las investigaciones de validación que se limitan sólo a ella. Al respecto Cronbach señaló en 1984: "toda validación es validación de constructo"⁽⁵⁾.
- La fase actual, que correspondería a la contextual o consecuencial, donde se amplía la fase anterior agregando el "uso propuesto". Su objetivo se relaciona con dotar de límites éticos a los métodos e instrumentos desarrollados. Esta revisión no modifica el concepto de validez (manteniendo una acepción unitaria) sino que exige delimitar los fundamentos teóricos a un contexto externo, como son las condiciones que rodean la situación en que se realiza la medición, entre otros aspectos^(4,6-9).

La ventaja substancial que subyace al concepto moderno de validez es que flexibiliza los procedimientos que permiten validar un instrumento; sin embargo es importante rescatar lo expresado por

Messik en relación a que la validez no puede descansar en un solo tipo de evidencia^(7,8). Por lo tanto, una investigación que busque validar un método no puede limitarse sólo a aportar validez de constructo, de criterio o de contenido.

Para la comprensión de la visión tripartita de validez, nos parece importante ahondar con mayores antecedentes una perspectiva conservadora en las ciencias médicas que se contempla en la mayoría de los procesos de validación de métodos de medición.

1. Validez de contenido: De acuerdo con lo señalado por Devlin y cols⁽¹⁰⁾, este concepto busca que la escala de medición presente las siguientes características:

- a) Induzca a un mínimo de respuestas sesgadas
- b) Sea fácil de entender e interpretar
- c) Sea fácil de administrar o aplicar en el trabajo de campo
- d) Posea capacidad de discriminar.

Convergen en ella la validez de contenido propiamente tal y la validez de fachada o aparente. La validez de contenido indica que la forma de interpretar los resultados parece ser razonable, según la visión de los expertos; por su parte, la validez de fachada se relaciona con la característica de calidad aparente, a juicio de expertos. Para la obtención de validez de contenido no se realiza ningún análisis empírico, sino que se recurre al juicio subjetivo de un panel de expertos en el área, que garantizan que la escala parece apropiada para el propósito de la medición. En 1954, Guilford llamó validación por presunción a esa forma de validación, en inglés *validity by assumption*⁽¹¹⁾. En imaginología, ésto puede ser ejemplificado en la determinación de que algunas imágenes son altamente sugerentes o patognomónicas de un tipo de patología, lo cual se asume como norma por la mayor parte de la comunidad científica. Por otro lado, si bien es cierto que la opinión de expertos es necesaria, aunque no libre de sesgos, es importante señalar que es rol de los autores el ponderar los costos en tiempo y recursos, además de los beneficios obtenidos al reemplazar, quitar o agregar elementos de juicio para la valoración de un determinado atributo; también es facultad de los mismos el considerar o no las sugerencias y correcciones provenientes del panel de expertos.

2. Validez de criterio: Para Streiner, la definición tradicional de validez de criterio es "la correlación de la escala con alguna otra medición del rasgo estudiado, idealmente con el patrón de oro o gold standard que ha sido aceptado en el área"⁽¹¹⁾. La validez de criterio, tal como fue mencionado en la fase operacional de validez, ha sido dividida en dos tipos: validez concurrente y validez predictiva e implica que la medición del instrumento se ajusta o sirve a un criterio externo. De la primera podemos decir que correlacionamos la

nueva escala con el criterio aceptado (patrón de oro) realizando una medición al mismo tiempo, es decir, el criterio se fija en el presente, como ha ocurrido con los estudios de validación de imágenes de las técnicas digitales actuales, que se contrastan con los resultados provenientes de técnicas convencionales y se analiza su concordancia^(12,13). En el segundo caso, la validez predictiva se denomina así porque medimos o clasificamos esperando un evento futuro, con el cual podemos establecer asociaciones. Ésto se podría representar, por ejemplo, como el valor del screening mamario (mamografía) en la sobrevida por carcinoma ductal *in situ*^(14,15).

En ocasiones, cuando se genera una escala para un constructo que no ha sido abordado en otro instrumento de medición y por lo tanto no existe gold standard, la validez de criterio simplemente no se contempla en el estudio de validación de la escala.

Validez de constructo: La validez de constructo es el principal tipo de validez, que consiste en un análisis de la significación de las puntuaciones de los instrumentos de medida, expresada en términos de los elementos que conformarán el diagnóstico imaginológico (tamaños, bordes, formas, densidad, etc).

Como primera consideración, la validación se lleva a cabo siempre por medio de investigaciones empíricas y estará representada por grados (dependiendo de la evidencia que se obtenga de una investigación empírica), más que por la presencia o ausencia total de ella. Como segunda consideración, no menos importante, tenemos que la validez no es propiedad de una escala, instrumento o método, sino que pertenece a las inferencias obtenidas a partir de la interpretación de los resultados de ellos⁽¹¹⁾, en otras palabras lo que se valida es el uso para el que se destina el instrumento o método de medición.

Aspectos operacionales para obtener evidencias de validez.

El proceso de validación de un instrumento, imagen o método, agrupa a un conjunto de análisis orientados a relacionar los resultados obtenidos del proceso de medición de elementos relacionados a un constructo con una interpretación coherente respecto de un uso y contexto definidos en términos de tiempo, espacio y población; de esta forma el cálculo de la validez va a estar condicionado a la contingencia en la que se ha generado el método de medición. Así, el procedimiento de validación asegura que las inferencias realizadas a partir del proceso de medición se encuentran avaladas por evidencias científicas que prometen representar la realidad.

Con el propósito de ilustrar los aspectos operacionales que deben considerarse en un estudio de validación, nos referiremos a aspectos esenciales obtenidos del libro Health Measurement Scales⁽¹¹⁾.

1. Validez de contenido: esta fuente de evidencia es

tipo subjetiva y reside en el criterio de expertos en el área en que se desarrolla el instrumento; la literatura tradicional describe un análisis cualitativo pero no cuantitativo. Es útil para el registro y sistematización de este proceso utilizar escalas tipo Likert. De esta forma, es esperable que si deseamos validar por ejemplo un nuevo método de diagnóstico para coleditiasis, deberemos indicar y describir el método en detalle, con las puntuaciones e inferencias provenientes de aquel o aquellos especialistas más destacados del área en nuestra realidad nacional o latinoamericana, por ejemplo; ellos demostrarán su acuerdo con el nuevo método, objetivándolo mediante una escala tipo Likert, antes mencionada. El acuerdo podrá ser presentado en términos de la media porcentual global entre los expertos.

2. Validez de criterio: dependiendo de la forma que tomen los resultados será la estrategia estadística que vamos a seleccionar para el estudio de validación por criterio. Por ejemplo, para la validez concurrente en la cual los resultados se han obtenido de la administración de dos instrumentos, léase uno gold standard y otro prueba índice donde los resultados sean expresados dicotómicamente, podemos recurrir al uso de una tabla de contingencia de 2x2 (Figura 1). Esta forma de análisis nos permite determinar índices de sensibilidad / especificidad y además, a partir de esta misma tabla, calcular un índice de correlación entre los instrumentos, como el coeficiente phi, cuya fórmula se describe en la Figura 2.

	Estado respecto a la enfermedad según el estándar de referencia	
	Presente	Ausente
Resultado de la prueba en estudio		
Positivo	a (enfermos con prueba +)	b (no enfermos con prueba +)
Negativo	c (enfermos con prueba -)	d (no enfermos con prueba -)

Figura 1. Tabla de contingencia para el estudio de validación por criterio con resultados dicotómicos.

$$\Phi = \frac{[BC - AD]}{\sqrt{[(A+B) (C+D) (A+D) (B+D)]}}$$

Figura 2. Ecuación de cálculo para el coeficiente phi.

MISCELÁNEOS

Si las mediciones corresponden a datos continuos, usaremos el coeficiente de correlación intra-clase. En este caso, esperamos encontrar una fuerte correlación entre los resultados de los dos instrumentos que están siendo confrontados.

Cuando se trata de establecer la validez predictiva del instrumento, se procederá de igual forma que lo anterior, teniendo en consideración las diferencias temporales entre la aplicación de un instrumento y de otro.

3. Validez de constructo: Para Nunnally y Bernstein⁽¹⁾, existen tres elementos importantes para la validación de un constructo:

- Especificar el dominio de variables observables relacionadas con el constructo
- Determinar el grado en que las variables observables tienden a medir lo mismo, varias cosas diferentes o muchas cosas diferentes, por medio de investigación empírica y análisis estadísticos.
- Realizar estudios subsecuentes y/o experimentos de diferencias individuales para determinar el grado en que las supuestas mediciones del constructo son consistentes con las «mejores conjeturas» acerca del constructo.

Desde el punto de vista operacional, debemos advertir que la determinación de la validez de constructo es la más compleja de establecer. Una de las estrategias para otorgar evidencia de validez de constructo es el uso de grupos extremos: uno donde se exprese altamente el rasgo o atributo a analizar y otro que idealmente no lo contenga; por ejemplo, estudiar una manifestación imaginológica determinada en sujetos sanos y enfermos. De esta forma, el grupo que posea el atributo (en este caso enfermedad) obtendrá un alto nivel en la escala de medición estudiada (presencia de manifestación imaginológica), mientras que el grupo sin el atributo o con poco de él puntuará muy bajo o cercano a 0 (ausencia de manifestación imaginológica).

Existe otra forma de establecer la validez de constructo, la matriz multirasgo – multimétodo⁽¹⁶⁾, técnica propuesta por Campbell y Fiske en 1959, que requiere para su aplicación:

- a) Seleccionar dos o más formas de medir el constructo en estudio (por ejemplo: determinar el número y tamaño de los cálculos biliares mediante ultrasonido y colangioRM).
- b) Identificar dos o más constructos que puedan medirse por los mismos métodos que el constructo de interés (ejemplo: número, tamaño y forma de los cálculos, compromiso de vía biliar extrahepática)
- c) Medir en una muestra todos los constructos con todos los métodos.

Una vez realizados los pasos anteriores, se

obtienen respuestas en las cuales se calculan las correlaciones entre todas las parejas de medidas posibles. Los coeficientes obtenidos se agrupan en una de las tres categorías siguientes:

1. *Coeficientes de fiabilidad*: son las correlaciones entre las medidas de un solo constructo con el mismo método. Teóricamente, estos coeficientes deben ser elevados.
2. *Coeficientes de validez convergente*: son las correlaciones entre las medidas de un solo constructo con métodos de medida diferentes. Estos coeficientes también debieran ser elevados, idealmente un poco menores que en el anterior (de fiabilidad).
3. *Coeficientes de validez discriminantes*: Son las correlaciones entre las medidas de diferentes constructos, obtenidas con el mismo método (coeficientes heterorasgo-monométodo), o las correlaciones entre las medidas de diferentes constructos con diferentes métodos (coeficientes heterorasgo-heterométodo). Estos deben ser los coeficientes más bajos de los 3 enunciados. Este método tiene la particularidad que además involucra aspectos de confiabilidad.

Comentario final

El conocimiento de la significancia de validez y algunas de las formas empíricas de evidenciarla, obliga al investigador del área imaginológica a hacerse parte en la crítica profunda de los métodos sobre los que basamos nuestros parámetros clasificatorios y de medición, lo que conlleva que la investigación de validación de métodos aumente sustancialmente, manteniéndose así aquellos que sean válidos y reconstruyendo o modificando los que no logren asemejarse a nuestra realidad.

Bibliografía

1. Nunnally JC, Bernstein IH. Teoría psicométrica. 2a ed. México: McGraw-Hill Interamericana, 1999; p. 179-184.
2. Greulich WW, Pyle SI. Radiographic atlas of skeletal development of the hand and wrist. 2nd ed. Stanford: Stanford University Press 1959; p. 61-183.
3. American Educational Research Association, American Psychological Association & National Council on Measurement in Education. Standards for educational and psychological testing. Washington: American Psychological Association 1999; p. 91-97.
4. Elousa OP. Sobre la validez de los tests. Psicothema 2003; 15: 315-321.
5. Cronbach LJ. Essentials of psychological testing, 4 ed. New York: Harper and Row; 1984; 269-275.
6. Messick S. Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. Am Psychol 1995; 50: 741-749
7. Messick S. Meaning and values in test validation: The science and ethics of assessment. Educ Res 1989;

- 18: 5-11.
8. Messick S. Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice EM:IP* 1995; 14: 5-8.
 9. Moss PA. The Role of Consequences in Validity Theory. *Educational Measurement: Issues and Practice EM:IP* 1998; 17: 6-12.
 10. Devlin S, Dong H, Brown M. Selecting a scale for measuring quality. *Marketing Research* 1993; 5: 12-17.
 11. Streiner DLN, Geoffrey R. *Health Measurement Scales: A Practical Guide to Their Development and Use*. Oxford: Oxford University Press 2003; p. 225-242.
 12. Hildebolt CF, Vannier MW, Shrout MK, Pilgram TK, Province M, Vahey EP, et al. Periodontal disease morbidity quantification. II. Validation of alveolar bone loss measurements and vertical defect diagnosis from digital bite-wing images. *J Periodontol* 1990; 61: 623-632.
 13. Sakai S, Soeda H, Takahashi N, Okafuji T, Yoshitake T, Yabuuchi H et al. Computer-aided nodule detection on digital chest radiography: validation test on consecutive T1 cases of resectable lung cancer. *J Digit Imaging* 2006; 19: 376-382.
 14. Geller BM, Kerlikowske K, Carney PA, Abraham LA, Yankaskas BC, Taplin SH, et al. Mammography surveillance following breast cancer. *Breast Cancer Res Treat* 2003; 81: 107-115.
 15. Uchida SM. Mamografía de Screening y Realidad Chilena. *Rev Chil Radiol* 2008; 14: 130-134.
 16. Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol Bull* 1959; 56: 81-105.