VALIDEZ DE CRITERIO DE INSTRUMENTOS DE MEDICIÓN BIOMÉDICA

CRITERION VALIDITY OF MEASUREMENT INSTRUMENTS USED IN BIOMEDICAL RESEARCH

Augusto Muñoz Caicedo*

RESUMEN

Objetivo: *Este artículo describe el proceso de validación de criterio de una prueba diagnóstica o de una escala de medición.* ***Método:*** *Se realizó un compendio de diferentes escritos consultados en libros y algunas bases de datos sobre este proceso.* ***Resultados:*** *para la validación de criterio de una prueba diagnóstica o de una escala de medición, se recomienda realizar cuatro pasos: seleccionar una prueba oro adecuada, determinar el tamaño de muestra, establecer la sensibilidad, especificidad y valores predictivos y, por último, determinar la concordancia entre las dos pruebas diagnósticas o las escalas.* ***Conclusiones:*** *La validación de criterio de una prueba diagnóstica o de una escala de medición, es un proceso sistemático que requiere de una adecuado desarrollo que permitan una adecuada interpretación de los resultados para evitar la difusión de instrumentos de medición poco confiables y con poca capacidad de discriminación diagnóstica.*

Palabras clave: Estudios de validación, Diagnóstico, Valor predictivo de las pruebas (DeCS).

ABSTRACT

Objective: *To describe the process of criterion validation of a diagnostic test or a measurement scale.* ***Methods:*** *I review a compendium of books and explored some databases to conduct a summary of the information.* ***Results:*** *For the process of criterion validation of a diagnostic test or a measurement scale, we always take four steps: select an appropriate gold standard, determine the sample size, set the sensitivity, specificity and predictive values. Finally, determine the agreement between two diagnostic tests or scales.* ***Conclusions:*** *The criterion validation of a diagnostic test or a measurement scale is a systematic process that requires proper development. In this way, the results of measurement instruments have a diagnostic capability.*

Keywords: *Validation Studies, Diagnosis, Predictive Value of Tests (MeSH).*

Historia del artículo:
Fecha de recepción: 04/05/2015
Fecha de aceptación: 08/09/2015

* Fonoaudiólogo, Magister en Salud Pública, Profesor Asociado de la Universidad del Cauca.
Correspondencia: Augusto Muñoz Caicedo. Dirección: calle 36 norte 4-114 casa 154, Popayán-Cauca.
Teléfono: 3167525189, correo electrónico: amunozc@unicauca.edu.co

INTRODUCCIÓN

Generalmente la medición del estado de salud se ha realizado desde la perspectiva biomédica del proceso salud-enfermedad mediante el uso de marcadores biológicos denominados desenlaces objetivos (1). Sin embargo, dado que la perspectiva meramente biológica resulta limitada, se considera la necesidad de la medición no biológica con indicadores subjetivos (2). En ese sentido, la medición de atributos subjetivos individuales o poblacionales se realiza mediante instrumentos, pruebas diagnósticas o escalas debidamente validadas.

En el área de la salud existen variedad de instrumentos de medición que, por lo general, han sido desarrollados en otros países que cuentan con lenguajes y culturas diferentes. El tener que aplicarlos en una población diferente, implica la necesidad de realizar el proceso de validación que se puede resumir en los siguientes pasos (3):

a. Adaptación cultural de la prueba o de la escala.

b. Validez de apariencia: la escala parece medirlo que verdaderamente debe medir

c. Validez de constructo: la escala no deja factores sin medir ni mide dominios que no son del síndrome.

d. Validez de criterio: la escala funciona de manera se mejante a otros instrumentos.

e. Confiabilidad test-retest o confiabilidad intereva luador: la escala funciona bien bajo diferentes condiciones de evaluación.

f. Sensibilidad al cambio: la escala detecta modificaciones de la realidad que mide.

g. Utilidad: es una escala fácil de aplicar y procesar.

Teniendo en cuenta la complejidad de cada uno de los pasos anteriormente descritos, este artículo se referirá específicamente a la validez de criterio.

Definición de la validez de Criterio

La validez de criterio es definida como el proceso mediante el cual se aplica a la escala a probar, un patrón de oro o gold estándar de referencia, y la prueba o escala evaluada, produce resultados que concuerdan con ese patrón de oro (4). Por ejemplo, es posible verificar si las respuestas en una escala que mide el dolor guardan una relación predecible con el dolor de una intensidad conocida (5).

Tipos de validez de Criterio

La validez de criterio se divide en validez concurrente y validez predictiva. La validez concurrente o medida externa de criterio, hace referencia a la correlación entre los resultados del nuevo instrumento y los resultados de una prueba o escala que ya ha sido probada. Las dos mediciones se realizan al mismo grupo de sujetos e, idealmente, al mismo tiempo, aunque también es aceptable realizarlas en una aproximación temporal (6). Por otro lado, la validez predictiva se puede entender como esa correlación entre el resultado del instrumento con un desenlace en el futuro. Implica probar los resultados obtenidos de un grupo de sujetos para un determinado constructo y luego compararlos con los resultados obtenidos en algún momento temporal posterior (7).

Pasos para la validación de Criterio

1. Seleccionar el Patrón de Oro o Gold Estándar

Es la primera condición que se debe evaluar para poder realizar la validez de criterio. La determinación del patrón de oro o gold estándar corresponde a la selección del método que ofrezca la mayor probabilidad de encontrar el valor real de un evento. En ese sentido, el patrón de oro más frecuentemente utilizado en medicina es el análisis de tejidos mediante autopsias o biopsias. Sin embargo, el avance de la tecnología en las ciencias de la salud, ha permitido que algunas pruebas, procedimientos o técnicas no invasivas utilizadas para el diagnóstico de algunas patologías, sean consideradas como patrón de oro, puesto que al aplicarlas, tienen una alta probabilidad de clasificar correctamente a un individuo sano (Sensibilidad de la prueba) y una alta probabilidad de clasificar correctamente a un individuo enfermo (Especificidad de la prueba) (8).

Por lo general, el patrón de oro es un examen invasivo que implica costos, tiempo y riesgos para el paciente, lo que constituye el principal motivo para la validación de pruebas o escalas con menor riego para el paciente, más rápido, menos costoso y más práctico. Sin embargo, para la medición de eventos en salud que no cuentan con una prueba oro como, por ejemplo, los sociales, psicológicos, cognitivos o de funciones superiores como el lenguaje, la atención o la memoria, algunos autores recomiendan realizar la comparación con métodos de apreciación clínica subjetiva, aunque en estos casos, se debe tener en cuenta que los valores de correlación entre las dos mediciones no suele ser muy alta (9).

2. Determinar el tamaño de la muestra

En la práctica clínica es muy frecuente comparar una técnica nueva con una ya establecida con el fin de determinar el grado de acuerdo entre las dos. Es importante resaltar el pa-

pel que juega el tamaño de muestra en este proceso y en ese sentido, algunos investigadores recomiendan tener presente si las variables a comparar son continuas, dicotómicas u ordinales. Para el caso de la proporción de desacuerdos entre las dos pruebas, se utilizan las formulas establecidas por autores como Machin (10); si la variable de interés es continua, recomiendan a Donner y Eliasziw (11); si es dicotómica a Flakck Afifi (12) y si es Ordinal a Cicchetti (13). Por otro lado, otros autores recomiendan que cuando se plantea un estudio de concordancia para determinar la intercambiabilidad de diferentes sistemas de medición, es necesario tener un tamaño de muestra que permita obtener el coeficiente de concordancia y correlación Pc (4). Al respecto, plantean utilizar la siguiente formula:

$$n = (Z_{1-\alpha/2} E / \pi P_c)^2$$

Fórmula para tamaños de muestra en estudios de concordancia.

Donde,

π= Diferencia porcentual esperada del verdadero valor del P_c.

E= Desviación estándar del P_c.

$Z_{1-\alpha/2}$= 100(1-α/2) percentil de la distribución normal estándar.

n= Numero de mediciones necesarias.

Por otra parte, el programa de libre difusión Epidat, permite calcular tamaños de muestra para estudios de concordancia cuando existen dos o más evaluadores.

3. Determinar la Sensibilidad, Especificidad y Valores predictivos de la prueba

Para establecer los valores de Sensibilidad, Especificidad y Valores Predictivos se hace necesario construir la tabla binaria, tetracórica o tabla 2x2. Una vez aplicada la prueba oro a los sujetos seleccionados, se aplica la prueba a validar y se clasifican según los nuevos resultados como positivos o negativos (4,5). Seguidamente, en la casilla (a) se coloca el número correspondiente a los sujetos diagnosticados como enfermos mediante las dos pruebas utilizadas (prueba oro y prueba o escala a validar). En la casilla (b), el número de sujetos que fueron diagnosticados como sanos con la prueba Oro y como enfermos con la prueba a validar. En la casilla (c), el número de sujetos diagnosticados como enfermos con la prueba Oro y sanos con la prueba a validar y por último, en la casilla (d) se coloca el número de sujetos que fueron diagnosticados como sanos mediante las dos pruebas utilizadas. Para algunos autores, el criterio de referencia de la presencia de una enfermedad no es solo otra prueba, sino también un periodo de seguimiento (5).

Figura 1. Tabla tetracórica o tabla de 2x2

	Prueba Oro o Gold Estándar	
Prueba a validar	a	b
	c	d

Donde,

a= Verdaderos positivos
b= Falsos positivos
c= Falsos negativos
d= Verdaderos negativos

Es importante tener presente que la prueba oro se debe ubicar en la parte superior de la tabla tetracórica y la prueba a validar en la parte lateral izquierda de la tabla (Figura 1).

Sensibilidad: la medición de la sensibilidad se define como la capacidad de la prueba para clasificar correctamente al sujeto enfermo como enfermo o como la probabilidad de tener un resultado positivo si se tiene la enfermedad. Puede definirse también en términos probabilísticos como la probabilidad de tener un resultado positivo dado que se está enfermo (4). Se establece utilizando la formula (a/a+c) y se considera una sensibilidad aceptable mayor a 70%.

Especificidad: la medición de la Especificidad es descrita como la capacidad de la prueba para clasificar correctamente al sujeto sano como sano o como la probabilidad de tener un resultado negativo si no se tiene la enfermedad. Igual que la sensibilidad, esta puede definirse en términos probabilísticos como la probabilidad de tener un resultado negativo dado que se está sano. Se establece utilizando la formula (d/b+d) y se considera una especificidad aceptable mayor a 70%.

Valor predictivo positivo: es definido como la probabilidad de tener la enfermedad dado que se tiene un resultado positivo; o en otras palabras, cual es la probabilidad de que el paciente detectado como verdadero enfermo mediante la sensibilidad, este realmente enfermo y no sea un falso positivo. Se establece utilizando la formula a/(a+b).

Valor predictivo negativo: se considera como la probabilidad de que el resultado negativo corresponda realmente a la ausencia de enfermedad; o interpretado como la probabilidad de que el sujeto detectado como sano mediante la especificidad, este realmente sano y no sea un falso negativo. La fórmula para establecerlo es d/(d+c)

Por lo general los valores de sensibilidad, especificidad y valores predictivos, se establecen en los casos cuando los resultados pueden clasificarse como variables dicotómicas. Sin embargo, para variables continuas se recomienda utilizar un punto de corte o calcular las razones de probabilidad (4).

4. Determinar la concordancia entre las dos pruebas

La concordancia evalúa si las diferentes técnicas utilizadas producen resultados similares cuando se aplican al mismo sujeto o grupo de sujetos y en forma simultanea o con mínimas diferencias de tiempo que garanticen ausencia de variabilidad. Se debe garantizar que las diferencias en los resultados no se deben a cambios fisiopatológicos en la variable medida (14).

Para determinar el grado de acuerdo se utiliza el estadístico Kappa y se interpreta como la proporción de concordancia más allá del azar (15). La fórmula para evaluar la concordancia entre variables nominales es:

$$K=((P(A)-P(E))/(1-P(E))$$

Fórmula para establecer el grado de acuerdo

Donde,

P(A)= Proporción de veces que los dos métodos o evaluadores concuerdan o están de acuerdo

P(E)= Proporción de veces que se espera que los dos métodos estén de acuerdo debida al azar únicamente.

Para la interpretación de los valores de Kappa se acepta la categorización realizada por Landis y Koch definida en la Tabla 1 (16).

Tabla 1. Interpretación de los valores del estadístico Kappa.

Valor de Kappa	Prueba Oro o Gold Estándar
<0	Pobre
0-0,20	Leve
0,21-0,40	Baja
0,41-0,60	Moderada
0,61-0,80	Buena
0,81-1,0	Casi perfecta

Por otro lado, para evaluar la concordancia entre dos variables continuas se utiliza el coeficiente de correlación de Pearson. En este caso, la hipótesis nula es considerada como la existencia de relación lineal entre los dos métodos. Lo anterior es anotado como una limitación del uso de r, dado que no responde a la pregunta de concordancia, limitándose a decir de manera simple que las dos técnicas están relacionadas linealmente (17).

Conclusiones

El proceso de validación de criterio de una prueba diagnóstica o de una escala de medición es un proceso sistemático que comprende cuatro pasos a saber: selección de una prueba oro adecuada, determinar el tamaño de la muestra, establecer la sensibilidad, especificidad y valores predictivos y por último, determinar la concordancia entre las dos pruebas.

Bibliografía

1. Luján-Tangarife, J.A, Cardona Arias, J.A. Construcción y validación de escalas de medición en salud: revisión de propiedades psicométricas. Med Pub Journals. Vol 11 No 3:1. Disponible en: http://www.archivosdemedicina.com/medicina-de-familia/construccin-y-validacin-de-escalasde-medicin-en-salud-revisin-depropiedades-psicomtricas.pdf
2. Grupo de la Organización Mundial de la Salud sobre la calidad de vida. Que calidad de vida? Foro Mundial de la Salud. Rev Inter Desar Sanit. 1996; 17: 385-7.
3. Sánchez R, Echeverry J. Validación de escalas de Medición en Salud. Revista de Salud Publica 6(3)302-318, 2004.
4. Ruiz A, Morillo L. Epidemiologia Clínica, Investigación Clínica aplicada. Editorial Panamericana, Pag 173.
5. Fletcher R, Fletcher S. Epidemiología Clínica. 4 edición, Barcelona, Editorial Wolters Kluver; 2007, p. 22
6. Streiner D, Norman GR, Health Measurement Scales. A Practical Guide to their development and use. Oxford University Press; 1995.
7. Shuttleworth, M. validez predictiva. 2010. Disponible en: https://explorable.com/es/validez-predictiva
8. Pita Fernández, S. Pertegas Diaz, S. Pruebas diagnósticas: Sensibilidad y Especificidad. Unidad de Epidemiología Clínica y Bioestadística. Cad Atención Primaria 2003; 10:120-124.
9. Ruiz, A. Uso de pruebas diagnósticas en medicina clínica. Capítulo 7; pag: 110.
10. Machin D, Campbell Mj, Fayers PM. Pinol APY. Sample Size Tables for Clinical Studies. Blackwell Science; 1997.
11. Walter SD, Eliasziw M. "Sample Size and optimal Designs for Reliability Studies" Statistics in Medicine 1998; 17: 101-10.
12. Flack VA, Afifi, Lachenbruch PA. "Sample Size Determinations for the two Rater Kappa statistic" Psychometrika 1988; 53(3):321-5.
13. Donner A, Eliasziw M. "A Goodnes of Fit Approach to Inference Procedures for the Kappa Statistics: Confidence Interval Construction, Significance Testing and Sample Size Estimation" Statistics in Medicine 1992; 11:1511-19.
14. Soledad, M. Estudios de Concordancia: Intercambiabilidad en sistemas de medición. Capítulo 17; pag: 294.
15. Ker M. Issues in the use for Kappa. Investigative Radiology 1991; 26:78-83.
16. Landis, JR. Koch GC. The measurement of observer agreement for categorical data. Biometrics 1977; 33:159-74.
17. Bland, JM. Altman, DG. Comparing two methods of clinical measurement: a personal history. Int J Epidemiol 1995;249 (suppl. 1):S7-S14.