

Resultados estadísticos vs juicio apropiado: Comentarios a la validación de un cuestionario sobre el conocimiento de la depresión mayor

Statistical results vs proper judgment: Comments on the validation of a questionnaire on knowledge of major depression

Sr. Editor:

Hace poco, se logró iniciar la validación de un instrumento que permite cuantificar el grado de conocimientos de los médicos sobre la depresión mayor (1). Esta importante contribución se opaca por serias malinterpretaciones hechas por los autores respecto a los procedimientos de confiabilidad y validez de contenido.

Específicamente, los autores creen acertado la existencia de la confiabilidad del ítem y la clasificación de los niveles de confiabilidad, y consideran favorables los resultados de inteligibilidad y pertinencia. Primero, lo que identifican como “confiabilidad alfa de cada pregunta” es “la confiabilidad alfa si se elimina la pregunta”, un procedimiento que automáticamente produce cualquier programa estadístico que contiene un módulo de análisis de la confiabilidad. Este método analiza el efecto de remover un ítem sobre el coeficiente alfa (2) y tiende a no ser recomendado para analizar la calidad del ítem, porque es un método esencialmente univariado y no considera el efecto multivariado sobre los demás ítems, y enfatiza únicamente la correlación con un criterio interno y no con criterios externos (2, 3), lo que tenderá a producir la conocida paradoja alta confiabilidad – baja validez. Debería ser claro que el coeficiente alfa se calcula para un conjunto de ítems que representan un puntaje unidimensional y no para cada ítem. Por lo tanto, la denominación “confiabilidad de la pregunta” es completamente inexacta.

Segundo, los niveles de confiabilidad adoptados por los autores no corresponden al proceso de obtener instrumentos con buenas propiedades métricas. Los autores consideran como aceptables índices de confiabilidad entre 0,6 y 0,8. Sin embargo, al verificar la fuente bibliográfica citada en su trabajo

(4), encontramos que se consideran como aceptables coeficientes por encima del 0,75. Esto último se acerca más a lo que la bibliografía especializada cita en relación a la interpretación del coeficiente alfa; solo por poner algunos ejemplos, Nunnally y Bernstein (5) declaran que un valor de 0,80 es recomendable, y para decisiones clínicas importantes 0,90 es ideal. Sattler (6) y Cicchetti (7) establecen como aceptable una magnitud desde 0,70 a 0,80, y buena entre 0,80 a 0,90. Entonces basados en la verificación de la fuente utilizada en el estudio y la revisión de la literatura especializada, los índices elegidos pueden considerarse como inaceptables para la estimación de una adecuada confiabilidad.

En último lugar, los autores declaran que los expertos que calificaron la validez de contenido mediante la inteligibilidad y pertinencia de los ítems, dieron una adecuada valoración, siendo los valores promedio 2,8 y 3,0, respectivamente. Tomando como referencia el rango de opciones utilizado en tal estudio (de 1 al 5), tales valores promedio están alrededor del descriptor “parcialmente inteligible” y “parcialmente pertinente”. Estos niveles podrían sugerir que el instrumento tiene un contenido parcialmente válido, lo cual difícilmente puede ser reemplazado por una afirmación que concluya que los ítems son adecuados. Incluso, los autores podrían estar de acuerdo que el mínimo puntaje recomendado para retener un ítem en esta etapa sería 4 (“inteligible” o “pertinente”).

Nuestro rol no es complementar la revisión por pares realizada para seleccionar y mejorar el artículo antes de su publicación, sino más bien resaltar los procedimientos metodológicos que fueron plausiblemente malinterpretados por los autores, dado que los lectores pueden tomarlos como respaldo para sus propios análisis.

César Merino-Soto

Magister en Psicología, Universidad de San Martín de Porres. Lima Perú.

Eduardo Manzanares-Medina

Bachiller en Psicología

Marisol Angulo-Ramos

Licenciada en Enfermería

REFERENCIAS BIBLIOGRÁFICAS

1. Ríos-Flores A, Leonardo-Olivera W, Ballena-López J, et al. Validación de un instrumento para medir el nivel de conocimiento sobre depresión mayor en médicos de atención primaria en Chiclayo, Perú. *Rev Med Hered.* 2013; 24:26-32.
2. Raykov T. Reliability if deleted, not “alpha if deleted”: Evaluation of scale reliability following component deletion. *Brit J Math Stat Psy.* 2007; 60: 201-16.
3. Raykov T. Alpha if item deleted: A note on criterion validity loss in scale revision if maximizing coefficient alpha. *Brit J Math Stat Psy.* 2008; 61: 275-85.
4. Hernandez R, Fernández C, Baptista P. Metodología de la investigación científica (4ª ed.) México: Mc Graw-Hill; 2006.
5. Nunnally J, Bernstein I. Teoría Psicométrica (3ra ed.). México, D.F: McGraw-Hill; 1995.
6. Sattler, JM. Assessment of children: Cognitive applications. (4a ed.). San Diego: Jerome M. Satler Publisher Inc.; 2001.
7. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assessment.* 1994; 6: 284-90.

Recibido: 12/08/2013

Comentarios de los autores

Author's response

En la elaboración de instrumentos psicométricos la evaluación de la confiabilidad y validez de los mismos es fundamental. En relación a la confiabilidad, dos son los procedimientos más utilizados: la medida de la consistencia interna (mediante el cálculo del coeficiente

de alfa de Cronbach o de Kuder y Richardson si la alternativa para cada ítem es dicotómica) y la medida de la estabilidad del instrumento en el tiempo (confiabilidad por test-retest) (1).

La literatura señala que el valor límite para aceptar un alfa de Cronbach es de 0,7 (2); el valor de 0,62 del instrumento en discusión, no cumpliría con este requisito. Sin embargo se debe recordar que el valor del coeficiente de confiabilidad puede aumentar en relación al número de entrevistados en la prueba piloto. En esta validación sólo se entrevistaron a 40 individuos. Según los registros del Colegio Regional VIII del Colegio Médico del Perú (Chiclayo) hasta Julio del año pasado había 1 320 médicos generales (3). Según Hernández Sampieri, el número ideal de participantes en un piloto sería entre el 10 al 20% del total de la muestra (1). El problema de accesibilidad a la mayoría de ellos hizo muy difícil alcanzar este número de participantes en el piloto.

En relación a los términos que aparecen en la tabla 1 del artículo: “confiabilidad de cada pregunta”, efectivamente, no son apropiados. Los términos adecuados son: “Valor del coeficiente de confiabilidad del instrumento si esta pregunta fuese eliminada”.

Se ha descrito que el coeficiente alfa de Cronbach puede sobreestimar o subestimar en algunas oportunidades la confiabilidad de un instrumento en validación, siendo usado en varias oportunidades en forma indiscriminada (4). Sin embargo su uso sigue siendo útil en los siguientes casos: Unidimensionalidad del test y valores de alfa mayores a 0,6 (4); éstas dos situaciones se presentan en el instrumento en validación. Asimismo la literatura reporta que en instrumentos con 20 o menos preguntas, el alfa de Cronbach como medida de confiabilidad, es fidedigno (5); el instrumento en discusión tiene 14 ítems y es unidimensional. La unidimensionalidad del instrumento (determinada mediante la imposibilidad de llevar a cabo un factorial por los valores de la prueba de esfericidad de Barlett y el coeficiente de Kaiser- Meir-Olkin), apoya éste concepto.

Cabe mencionar que en los software más utilizados actualmente en bioestadística: STATA v 11 y SPSS v 19, el cálculo de la confiabilidad es determinada fundamentalmente por el coeficiente de alfa de Cronbach.

Actualmente la tendencia es utilizar técnicas estadísticas más sofisticadas para determinar la