

- Journal of the Stat Assoc. 2001; 96(453): 210-219.
5. Hageman WJ, Arrindell WA. Establishing clinically significant change: Increment of precision and the distinction between individual and group level of analysis. Behav Research Ther. 1999; 1: 1169-93.

Recibido: 25/06/2013

Comentarios de los autores

Author's response

El concepto de fiabilidad es una manera de expresar o reflejar la cantidad de error, tanto aleatorio y sistemático, inherente a cualquier medición, por lo tanto la confiabilidad de un instrumento es el grado en que esta es reproducible, con casi el mismo valor cada vez que se la mida. Por lo tanto la confiabilidad o precisión de una medición es una función del error aleatorio (variabilidad por azar), cuanto mayor sea el error, menos precisa será la medición.

Existen tres fuentes principales de error aleatorio cuando se hacen mediciones: la variabilidad del observador, la variabilidad del instrumento y la variabilidad biológica intrínseca del participante. En muchos cuestionarios auto aplicados de función psicológica a menudo no hay observadores implicados en las mediciones, pero es importante conocer la confiabilidad del cuestionario o escala de medición. El enfoque usual es administrar en dos ocasiones separadas por un intervalo de tiempo razonable, donde se asume que el proceso es improbable que cambie. En este enfoque llamado confiabilidad test-retest se selecciona un intervalo de tiempo apropiado para aplicar el cuestionario que no sea demasiado largo para evitar cambios muy grandes, ni muy corto que permita el aprendizaje. Los expertos opinan que el intervalo de tiempo apropiado varía de horas a años, dependiendo de la tarea, pero un test-retest con un intervalo de 2 a 14 días es lo que se recomienda (1).

Cuando se realizó la validación psicométrica del instrumento Índice Internacional de Función Eréctil (IIFE) en pacientes con Disfunción Eréctil (DE) (2), la confiabilidad se evaluó midiendo la fiabilidad

test-retest, la consistencia interna y la sensibilidad al cambio.

La fiabilidad test-retest se evaluó aplicando el IIFE a los pacientes con DE en dos periodos con intervalo de una semana. Se empleó el coeficiente de correlación de Spearman para correlacionar el puntaje total y el puntaje por dominios obtenido en la primera evaluación con la segunda evaluación, y para evaluar la concordancia entre los resultados según la severidad, se empleó el estadístico Kappa.

Para la fiabilidad test retest, el análisis estadístico de las repuestas dadas en la primera prueba respecto de la segunda en el dominio función eréctil, arroja un $r_s = 0,58$ ($p < 0,001$), lo que indica una correlación moderada entre ambos resultados. Correlaciones semejantes se observan entre los demás dominios. La correlación entre el puntaje total de la primera evaluación y de la segunda da un $r_s = 0,8$ ($p < 0,001$), por lo que hay una correlación alta. También se encontró una concordancia débil ($Kappa = 0,29$, $p = 0,03$) pero estadísticamente significativa entre el resultado de severidad del primer examen respecto del segundo.

En la carta dirigida a los editores de la Revista Médica Herediana (RMH), se estima que los resultados de la fiabilidad test-retest están interpretados inapropiadamente, sustentando la observación, que los resultados se obtuvieron en el curso de un tratamiento para DE, y que en el intervalo de una semana se esperaba que ocurran cambios en la función eréctil (FE) como efecto exitoso de la terapia, mostrando correlaciones bajas o moderadas debido a la inestabilidad del constructo en la situación experimental inducida.

En el estudio de validación del IIEF, la fiabilidad test-retest se evaluó aplicando el IIEF a 35 pacientes con DE sin tratamiento en dos ocasiones, separadas por un intervalo de 5 ± 2 días (Datos no mostrados en el artículo), por lo que nuestras interpretaciones son correctas.

Según los autores de la carta, también sugieren que se pudo evaluar la estabilidad del test en los controles, sin la intervención, lo que nos parece apropiado, cuyos resultados estimamos hubieran mostrado los mismos resultados que el grupo de pacientes con DE.

Los resultados del test-retest en nuestro estudio son mostrados por dominios y mediante el puntaje total. Para mostrar los resultados se ha usado el coeficiente

de correlación de rangos de Spearman, ya que se han correlacionado los puntajes o scores del IIEF que fluctúan desde 0 a 30 puntos y constituye una variable ordinal. El coeficiente de correlación de rangos de Spearman es un método no paramétrico y no asume una distribución normal. Es una medida de asociación lineal que utiliza rangos, o números ordinales y compara dichos rangos. Como el coeficiente de correlación de Pearson es influenciado por varianzas desiguales, no normalidad, no linealidad y es muy sensible a valores atípicos, el coeficiente de correlación de Spearman es una medida de asociación más firme (3,4).

Un diseño común para la investigación del error de medición es tomar pares de mediciones sobre un grupo de sujetos. Cuando tenemos pares de observaciones, es natural *plotear* (trazar) una medida contra el otro. El diagrama de dispersión resultante puede tentarnos para calcular un coeficiente de correlación entre la primera y la segunda medición. Según Bland and col; hay dificultades en la interpretación de este coeficiente de correlación. En general, la correlación entre las mediciones repetidas dependerá de la variabilidad entre sujetos. Muestras que contengan sujetos que difieren en gran medida van a producir mayores coeficientes de correlación que muestras que contienen sujetos similares (5).

El coeficiente de correlación depende por tanto de la forma en que la muestra se elige, y que sólo tiene sentido para la población de la que los sujetos de estudio pueden considerarse como una muestra aleatoria. Si seleccionamos sujetos que dan una amplia gama de medidas, que es el enfoque natural cuando investigamos error de medición, éste inflará el coeficiente de correlación.

Otro problema con el uso del coeficiente de correlación entre la primera y segunda medición es que no hay razón para suponer que el orden o ranking de las mediciones sea importante. Si el orden o ranking de las mediciones son importantes no debería haber observaciones repetidas de la misma cosa. Podríamos invertir el orden de cualquiera de los pares o parejas de observaciones y obtener un valor ligeramente diferente del coeficiente de correlación entre las mediciones repetidas. Por ejemplo, invirtiendo el orden de igual número de sujetos obtendríamos un valor de $r = 0,80$ en lugar de un $r = 0,77$. El coeficiente de correlación intra-clase evita este problema, estimando una correlación promedio entre todas las posibles parejas ordenadas. Para la mayoría de las aplicaciones, sin embargo, Bland prefiere la medida de la desviación

estándar dentro de los sujetos como un índice de error de medición, ya que tiene una interpretación más directa que se puede aplicar a las mediciones individuales (5).

Las observaciones de Bland y de Streiner sobre las limitaciones de los coeficientes de correlación de Pearson, se han efectuado cuando se han evaluado la variabilidad inter-observadores e intra-observadores. Ellos prefieren el uso del coeficiente de correlación intraclase (CCI) y la medida de la desviación estándar dentro de los sujetos como un índice de error de medición (1,6). Pero estas observaciones no se han hecho cuando se han efectuado mediciones de la fiabilidad test-retest, con variables ordinales que no cumplen los supuestos de aleatoriedad y de normalidad. Por lo tanto consideramos que el coeficiente de correlación de Spearman continua siendo una medida de la confiabilidad más firme para evaluar el test-retest para las condiciones de nuestro estudio.

César A. Loza

Profesor Asociado, Universidad Peruana Cayetano
Heredia. Lima, Perú.

Correspondencia:

César A. Loza Munarriz
Av. Honorio Delgado 430 Urb. Ingeniería, San Mantín
de Porres. Lima 31, Perú.
Correo electrónico: cesar.loza@upch.pe

REFERENCIAS BIBLIOGRÁFICAS

1. David L. Streiner, Geoffrey R, Norman. Oxford University Press. Health Measurement Scales. A Practical Guide to Their Development and Use. Third Edition. Oxford, Great Britain. 2003. p. 127 – 152.
2. Zegarra L, Loza C, Pérez V. Validación psicométrica del Instrumento Índice Internacional de Función Eréctil en pacientes con disfunción eréctil en Perú. Rev Perú Med Exp Salud Pública. 2011; 28 (3):477 - 83.
3. Jennifer Peat, Belinda Barton. BMJ Books. Medical Statistics. A Guide to Data Analysis and Critical Appraisal. First edition. 2005. p. 156 – 201.
4. Marcello Pagano, Kimberlee Gauvreau. Thomson Learning. Principles of Biostatistics. Second Edition. 2000. p. 398 – 414.
5. Bland JM, Altman DG. Statistics Notes: Measurement error and correlation coefficients. BMJ. 1996; 313:41.
6. Bland JM, Altman DG. Statistics Notes: Measurement error. BMJ. 1996; 312:1654.