

# Modelos MIMIC: de la neurociencia a las ciencias de la salud

María F. Grande Ratti<sup>1</sup>, Romina Y. Pérez Manelli<sup>1</sup>, Fernando R. Vázquez Peña<sup>2</sup> y Eulogio Cordón Pozo<sup>3</sup>

1. Área de Investigación en Medicina Interna, Servicio de Clínica Médica, Hospital Italiano de Buenos Aires. Buenos Aires, Argentina

2. Servicio de Medicina Familiar y Comunitaria, Hospital Italiano de Buenos Aires. Buenos Aires, Argentina

3. Departamento de Organización de Empresas II. Universidad de Granada. España

## RESUMEN

En medicina, muchas veces se encuentran situaciones donde la regresión (lineal o logística) no es viable, por contar con constructos latentes o factores que incluyen diversas variables observadas, ya que tratan de reflejar o estimar fenómenos complejos multidimensionales. Esta limitación dio lugar a la búsqueda de metodologías estadísticas alternativas, como un subtipo de Modelos de Ecuaciones Estructurales denominados modelos MIMIC (del inglés, por *Multiple Indicators Multiple Causes*), utilizado originalmente en neurociencias. En esta entrega se explica qué son estos modelos, para qué sirven, así como se mencionan los programas para su aplicación, y la interpretación de resultados. Se ejemplifica con un estudio sobre la calidad de vida de personas que recibieron un trasplante, destacando la representación gráfica y la importancia de la parsimonia en los modelos. Estos modelos permiten incluir factores y variables sueltas en simultáneo, resolviendo relaciones causales complejas, por lo cual resultan herramientas útiles en la investigación médica.

**Palabras clave:** modelización de ecuaciones estructurales, modelos estadísticos, método estadístico, mecanismos de evaluación de la atención sanitaria.

## MIMIC Models: from Neuroscience to Health Sciences

### ABSTRACT

In medicine, situations are often encountered where regression (linear or logistic) is not feasible, due to latent constructs or factors that include several observed variables, since they try to reflect or estimate complex multidimensional phenomena. This limitation led to the search for alternative statistical methodologies, such as a subtype of Structural Equation Models called MIMIC (Multiple Indicators Multiple Causes) models, originally used in neurosciences. In this issue we explain what these models are, what they are used for, as well as the programs for their application, and the interpretation of results. It is exemplified with a study on the quality of life of people who received a transplant, highlighting the graphical representation and the importance of parsimony in the models. These models allow the simultaneous inclusion of factors and single variables, solving complex causal relationships, making them useful tools in medical research.

**Keywords:** structural equation modeling, models, statistical, statistical method, health care evaluation mechanisms.

---

Autor para correspondencia: [maría.grande@hospitalitaliano.org.ar](mailto:maría.grande@hospitalitaliano.org.ar), Grande Ratti MF.

Recibido: 16/07/24 Aceptado: 27/08/24 En línea: 10/10/24

DOI: <http://doi.org/10.51987/revhospitalbaire.v44i3.344>

**Cómo citar:** Grande Ratti MF, Pérez Manelli RY, Vázquez Peña FR, Cordón Pozo E. Modelos MIMIC: de la neurociencia a las ciencias de la salud. Rev. Hosp. Ital. B. Aires. 2024;44(4):e0000344

## INTRODUCCIÓN

La neurociencia cognitiva ha inspirado una serie de avances metodológicos en salud porque utiliza técnicas populares para resumir datos de comportamiento que incluyen puntuaciones, en donde la dimensionalidad de los ítems y la calidad de la información a menudo quedan sin explorar en profundidad<sup>1</sup>. Por ejemplo, el instrumento PHQ-9 es ampliamente utilizado para diagnosticar y determinar la gravedad de la depresión en atención primaria<sup>2</sup>. Sin embargo, la puntuación total se obtiene sumando el valor de cada una de las respuestas; y luego existen puntos de corte que representan depresión leve, moderada, moderadamente grave y grave (con 5, 10, 15 y 20 puntos, respectivamente). En este escenario, se recolectan 9 ítems independientes, cuyo contenido permite construir una variable numérica continua global, que a su vez se interpreta y se informa como categórica múltiple y ordinal. Se pasa de la granularidad a constructos, a través de transformaciones y operacionalización de variables.

Antes de seguir avanzando, definiremos “variables” y “factores”, debido a que tienen funciones clave en la estructura y el análisis. Los factores se refieren a los constructos latentes que no son directamente observables, pero que en la práctica se infieren a partir de las variables. Es el dominio que las integra, representan los conceptos teóricos subyacentes, y, por consiguiente, permiten reducir la cantidad de datos para analizar. En cambio, las variables son las mediciones concretas y observables que se registran directamente y que se utilizan para medir esos constructos latentes. Clásicamente se pueden dividir en dos tipos: dependientes (Indicadores) porque indican el estado del factor latente; e independientes (Causas) porque causan el factor latente. Por ejemplo, si el factor fuera “salud mental”, los indicadores podrían ser respuestas a encuestas sobre ansiedad y depresión, mientras que las causas podrían incluir variables como nivel de estrés y situación económica. Por ende, los factores son influenciados por las causas, y medidos a través de los indicadores.

Adicionalmente, repasaremos la clasificación de las variables según su naturaleza (o tipo de dato). Pueden ser: A) Cualitativas o Categóricas, aquellas que describen cualidades o características, y Nominales (no tienen un orden intrínseco; p. ej., hombre/mujer) u Ordinales (tienen jerarquía; p. ej., nivel educativo); B) Cuantitativas o Numéricas, aquellas que representan cantidades, y pueden ser Discretas (toman valores enteros y finitos; p. ej., número de hijos) o Continuas (pueden tomar cualquier valor en un rango; p. ej., edad, peso o glucemia).

En medicina es muy frecuente el uso de variables dicotómicas, con aproximación demasiado simplista (p. ej., hospitalización sí/no). Recordemos que las variables dicotómicas (o binarias) solo pueden tomar dos valores distintos (p. ej., enfermedad presente o ausente). Por ende, en más de una ocasión nos encontramos con situaciones en las que quisiéramos diseñar un modelo que permita incluir conjuntamente factores (compuestos por diversas variables) y variables sueltas (p. ej., género, edad). Para estos casos, pensamos en utilizar herramientas como la

regresión (logística o lineal), pero no podemos hacerlo por los factores; o pensamos en elaborar un Modelo de Ecuaciones Estructurales clásico (MEE es una técnica estadística multivariante para probar y estimar relaciones causales a partir de datos y suposiciones cualitativas sobre la causalidad), pero tampoco podemos hacerlo por las variables sueltas.

Los modelos lineales o logísticos no siempre son apropiados en todos los escenarios, debido a ciertas limitaciones inherentes a estas técnicas y los supuestos subyacentes. Aunque ambos métodos se utilizan para hacer predicciones, difieren en el tipo de variable dependiente que modelan. La regresión lineal se utiliza para modelar la relación entre una variable dependiente continua y una o más variables independientes (que pueden ser continuas o categóricas). La regresión logística se utiliza para modelar la probabilidad de ocurrencia de un evento binario en función de una o más variables independientes. Sin embargo, ambas utilizan una variable dependiente y una o varias variables independientes. Son todas variables observadas y no trabajan con factores o variables latentes. No permiten: A) que una variable sea dependiente e independiente a la vez, en el mismo modelo; ni B) hacer incidir una variable sobre otra, y a la vez esta última sobre la primera (modelos denominados “no recursivos”). Por otro lado, los MEE tradicionales tampoco permiten trabajar con variables sueltas (no factorizadas).

La alternativa que hoy en día permite responder a esta complejidad es usar los Modelos MIMIC (del inglés *Multiple Indicators Multiple Causes*). Aunque se han aplicado con éxito en investigaciones de salud mental<sup>3</sup> y geriátricas<sup>4</sup>, hay muchas otras secciones donde aún no se han aplicado, a pesar de tratarse de un enfoque metodológico conocido desde hace tiempo<sup>5</sup>. Sin embargo, la posibilidad de utilizarse con variables ordinales, con el programa Lisrel, fue conocida fundamentalmente a partir de las conferencias de Jöreskog, dadas en Alemania durante el año 2001. Tomando este evento como inicial, la evolución de los estimadores por distribución asintótica libre (del inglés *Asymptotically Distribution-Free*) ha sido crucial. Estos no necesitan el supuesto de normalidad multivariada y, por consiguiente, permiten trabajar con variables ordinales, que habitualmente no la presentan. El primero fue el *Weighted Least Squares* (WLS), en Lisrel; luego surge el *Diagonally Weighted Least Squares* (DWLS)<sup>6</sup>, permitiendo trabajar incluso con menores tamaños muestrales.

Actualmente encontramos, en Mplus y R Lavaan, el WLSMV o el DWLS Robusto<sup>7</sup>, que consiguen un desempeño aceptable, en todas las condiciones, incluso con tamaños muestrales pequeños, con un mínimo de 100 observaciones<sup>8</sup>.

### ¿Qué son y para qué sirven los modelos MIMIC?

Se trata de MEE que permiten trabajar con factores (variables latentes) y variables sueltas (no factorizadas), brindando la posibilidad de hacer regresiones múltiples tanto con factores como con variables, en simultáneo. Es una técnica que implica el uso de variables latentes que son predichas por variables observadas<sup>9</sup>.

La idea principal detrás de los modelos MIMIC es poder incorporar, en el modelo, variables observadas, como dependientes o independientes, de las variables latentes. Para entender esto, primero deberíamos recordar las definiciones de variables latentes, es decir, aquellas no observadas directamente, pero se asume que influyen en las variables observadas, y de variables observadas, aquellas medidas que se registran directamente. Por ejemplo, un estudio menciona que el constructo “dislipidemia” está compuesto sobre diferentes mediciones séricas de colesterol<sup>10</sup> (Fig. 1).

### ¿Por qué es importante trabajar con factores?

Cuando queremos estimar constructos o indicadores complejos, resultaría insuficiente hacerlo con una sola variable. Por ejemplo ¿podemos estudiar “Síndrome metabólico”, “Calidad de vida”, o “Hábitos saludables” con una sola variable? En la práctica clínica es común abordar estos fenómenos utilizando un enfoque multidimensional, que implica la medición y la consideración de diversas variables (factores), debido a que son conceptos complejos.

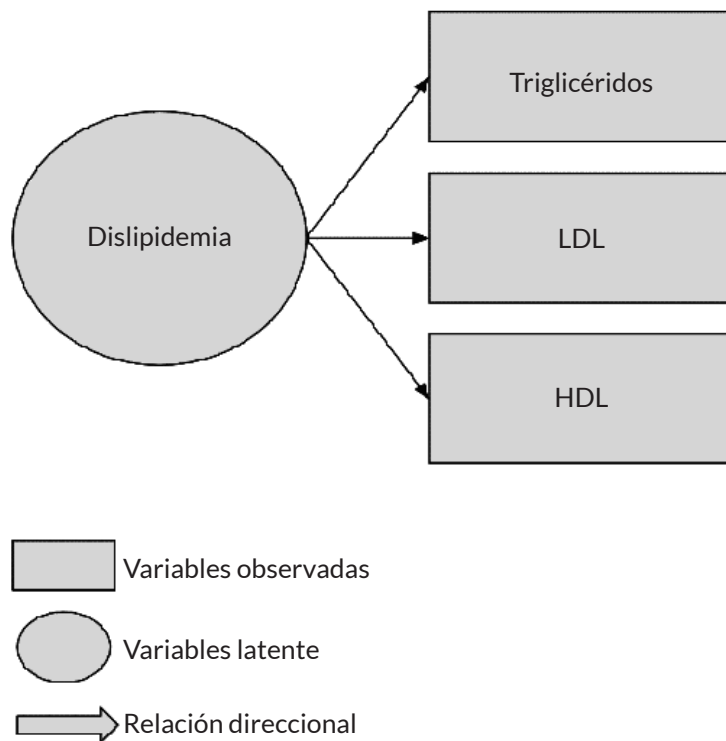
A veces, también queremos darle más estabilidad al modelo y, por lo tanto, preferimos el uso de factores. Por ejemplo, si estamos usando como variable independiente “nivel de colesterol total”, sería más estable utilizar “perfil lipídico”, incluyendo las mediciones de LDL y triglicéridos, si esto fuera apropiado desde lo conceptual.

Sin embargo, en otras disciplinas (como psicología o administración de empresas), es más común encontrar estudios que solo contienen factores y, por lo tanto, usan MEE clásicos. Pero, en medicina, casi siempre tenemos variables difíciles de factorizar, al menos en la mayoría de los estudios clínicos, como por ejemplo edad o género, o dicotómicas (p. ej., falleció o no).

### Dibujo en un ejemplo real

Un estudio de corte transversal incluyó la percepción autoinformada de 361 personas que recibieron un trasplante de órgano sólido, evaluando calidad de vida a través de herramientas como EuroQol (instrumento de medición de la calidad de vida relacionada con la salud, en el cual cada persona valora su estado de salud) y SF-36v2 (otro instrumento de medición de la calidad de vida relacionada con la salud, en el que, también, cada persona valora su propio estado de salud). Asimismo incluyó variables observables no factorizadas, como edad, género, nivel educativo, cantidad de trasplantes y tiempo en años del último trasplante.

El primer instrumento incluye una Escala Visual Analógica utilizada en la evaluación de la calidad de vida relacionada con la salud, herramienta que permite a los encuestados expresar su percepción subjetiva sobre su salud mediante la colocación de una marca en una línea horizontal, que va de 0 (representando la peor salud imaginable) a 100 (representando la mejor salud imaginable).



**Figura 1.** Constructo latente (dislipidemia) compuesto sobre variables observadas (mediciones de colesterol). LDL: lipoproteína de baja densidad. HDL: lipoproteínas de alta densidad. Fuente: elaboración propia, adaptación de original<sup>10</sup>.

El segundo instrumento consta de 36 preguntas que se agrupan en 8 dimensiones de salud: salud general, vitalidad, función física, papel físico, dolor corporal, salud mental, papel emocional y función social.

El indispensable primer paso es pensar y hacer un dibujo, que generalmente se refiere a especificar la estructura del modelo, incluyendo las relaciones entre las variables latentes y observadas, así como las relaciones causales entre las variables. Esto implica esquematizar gráficamente el modelo, buscando entender la relación entre una variable latente y sus indicadores observados, considerando también la influencia de variables externas o covariables. En la representación gráfica (Fig. 2), las variables latentes se suelen representar como óvalos o círculos (en color azul), mientras que las variables observadas (indicadores) se representan como rectángulos (en color verde); las flechas, entre ellas, indican las relaciones causales.

### Modelo parsimonioso

Este término se refiere a la preferencia por aquellos modelos más simples y concisos cuando se explican los datos, sin sacrificar en gran medida la capacidad de ajuste, buscando evitar la complejidad innecesaria<sup>11</sup>. Es decir, se busca un modelo que logre el equilibrio entre ajustarse bien a los datos y ser lo más simple posible.

### Programas estadísticos

Los modelos MIMIC se pueden estimar en muchos paquetes estadísticos, por ejemplo R Lavaan, Stata, MPlus y SAS, donde los resultados obtenidos son altamente convergentes<sup>12</sup>. La especificación y pruebas relativamente simples de los modelos MIMIC se pueden dividir en tres pasos secuenciales:

(1) Dibujar el modelo MIMIC, que permite visualizar de manera clara la estructura, y cómo las variables

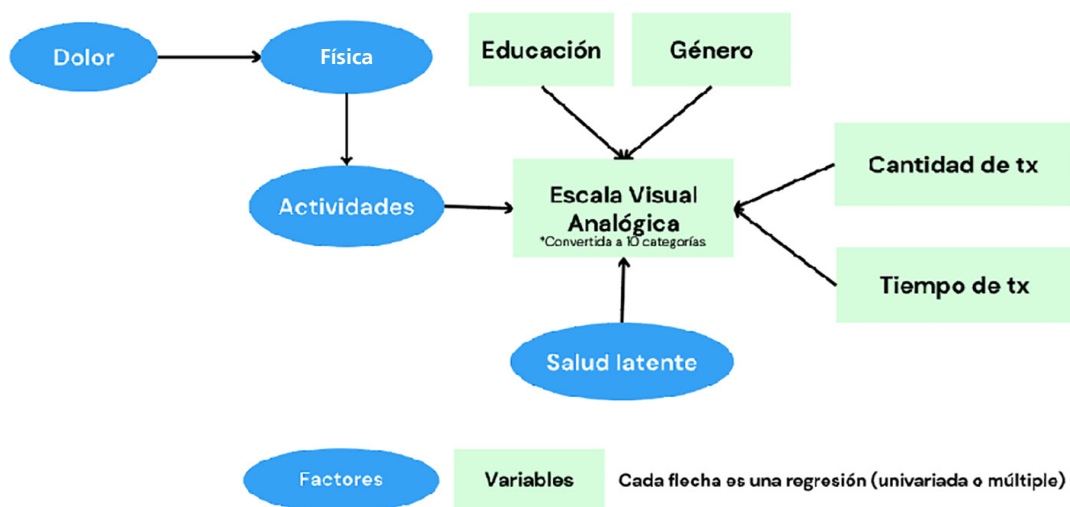
latentes y los indicadores observados están interrelacionados, lo que facilita la comprensión del modelo y su interpretación.

- (2) Desarrollar un modelo parsimonioso, que se ajuste bien para la muestra en su conjunto, de modo tal que los índices sean significativos y con magnitud suficiente.
- (3) Interpretar y generalizar los resultados con precaución, teniendo en cuenta las limitaciones y los supuestos del modelo. Se trata de evaluar lo bien que nuestros datos se han “ajustado” al modelo propuesto, información provista por la prueba de chi-cuadrado, los índices *ad hoc* y la significancia de los parámetros y/o coeficientes estimados. Se conoce que, en este tipo de estudios con regresiones, conviene utilizar los coeficientes estandarizados para no tener inconvenientes de interpretación con el tipo de unidades que usamos para categorizar las variables (Fig. 3). En este caso, los coeficientes estandarizados de los modelos MIMIC se interpretan como los estimadores beta de las regresiones lineales.

### Ventajas de los MEE

Usando modelos MIMIC tendremos unas ventajas adicionales que son poco conocidas en el ámbito médico, aunque frecuentemente usadas en otras disciplinas:

- Las variables pueden ser independientes y dependientes a la vez. Es decir, no hay variables exclusivamente independientes y exclusivamente dependientes. Una variable o factor A puede ser independiente en su relación con un factor B, y dependiente a su vez de una variable Z<sup>13</sup>.
- Existen también los modelos no recursivos. La variable A incide sobre la variable B, y a su vez la variable B incide sobre la variable A. Estos modelos, por ejemplo, se usan en violencia intrafamiliar. En



**Figura 2.** Una de las representaciones gráficas de un proyecto real, basado en calidad de vida autoinformada por personas que han recibido un trasplante (tx). Fuente: elaboración propia.

Regressions:						
	Estimate	Std.Err	z-value	P (> z )	Std.lv	Std.all
salud ~						
comosiento	-0.099	0.083	-1.192	0.233	-0.071	-0.070
actividades	0.157	0.078	2.015	0.044	0.122	0.121
saludlatente	0.833	0.098	8.460	0.000	0.517	0.510
x11	0.055	0.021	2.588	0.010	0.055	0.149
x5	-0.004	0.005	-0.890	0.373	-0.004	-0.052
x3	0.059	0.117	0.503	0.615	0.059	0.029
actividades ~						
emocional	0.041	0.053	0.774	0.439	0.050	0.050
física	0.649	0.054	12.070	0.000	0.744	0.744
emocional ~						
dolor	-0.698	0.048	-14.428	0.000	-0.627	-0.627
física ~						
dolor	-0.990	0.042	-23.417	0.000	-0.944	-0.944

**Figura 3.** Uno de los *resultados* del modelo MIMIC (en el caso real), utilizando R Lavaan. Nos interesan fundamentalmente los valores que arrojan las columnas "Std. all" (coeficientes estandarizados) y "P (>|z|)" (valor de p). x11: estudios formales de los pacientes, en años; x5: edad de los pacientes, en años; x3: género de los pacientes.

cambio, el clásico modelo recursivo es aquel en el cual las relaciones causales entre las variables son unidireccionales y acíclicas, es decir, no hay retroalimentación entre las variables<sup>14,15</sup>.

## CONCLUSIONES

Los modelos MIMIC son herramientas analíticas útiles para comprender y modelar relaciones complejas entre variables latentes y observadas, considerando la posibilidad de múltiples factores causales.

**Financiamiento:** el presente proyecto no obtuvo financiamiento de ninguna índole.

**Conflictos de intereses:** los autores declaran no poseer conflictos de intereses.

**Contribuciones de los autores:** Conceptualización, Curación de datos, Análisis formal, Investigación, Software: MFG, RYPM, FRVP. Metodología, Administración del proyecto: MFG y FRVP. Supervisión, Validación, Visualización, Redacción –borrador original, Redacción– revisión y edición: MFG, RYPM, FRVP, ECP.

## REFERENCIAS

- Rosen AFG, Auger E, Woodruff N, et al. The multiple indicator multiple cause model for cognitive neuroscience: an analytic tool which emphasizes the behavior in brain-behavior relationships. *Front Psychol.* 2022;13:943613. <https://doi.org/10.3389/fpsyg.2022.943613>.
- Costantini L, Pasquarella C, Odone A, et al. Screening for depression in primary care with Patient Health Questionnaire-9 (PHQ-9): a systematic review. *J Affect Disord.* 2021;279:473-483. <https://doi.org/10.1016/j.jad.2020.09.131>.
- Liu T, Peng MM, Wong FHC, et al. Differential associations between depressive symptom-domains with anxiety, loneliness, and cognition in a sample of community older chinese adults: a multiple indicators multiple causes approach. *Innov Aging.* 2023;7(7):igad075. <https://doi.org/10.1093/geroni/igad075>.
- Rubio L, Dumitrache CG, García AJ, et al. Coping strategies in Spanish older adults: a MIMIC model of socio-demographic characteristics and activity level. *Aging Ment Health.* 2018;22(2):226-232. <https://doi.org/10.1080/13607863.2016.1247416>.
- Jöreskog KG, Goldberger AS. Estimation of a model with multiple indicators and multiple causes of a single latent variable. *J Am Stat Assoc.* 1975;70(351):631-639. <https://doi.org/10.1080/01621459.1975.10482485>.
- Muthén BO. Goodness of fit with categorical and other nonnormal variables. En: Bollen KA, Long JS, editors. *Testing structural equation models.* Newbury Park, CA: Sage; 1993. p. 205-234.
- Muthén B, du Toit SH, Spisic D. Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes: technical report [Internet]. Los Angeles, CA: University of California; 1997 [consulta 2024 mar 01]. Disponible en [https://www.statmodel.com/download/Article\\_075.pdf](https://www.statmodel.com/download/Article_075.pdf).
- Flora DB, Curran PJ. An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychol Methods.* 2004;9(4):466-491. <https://doi.org/10.1037/1082-989X.9.4.466>.
- Schumacker RE, Lomax RG. *A beginner's guide to structural equation modeling.* 3rd ed. New York, NY: Routledge; 2012.
- Tekwe CD, Carter RL, Cullings HM, et al. Multiple indicators, multiple causes measurement error models. *Stat Med.* 2014;33(25):4469-4481. <https://doi.org/10.1002/sim.6243>.

11. Espinoza-Bello M, Galindo-Vázquez Ó, Jiménez-Ríos MA, et al. Propiedades psicométricas de la Escala de Evaluación Funcional para el Tratamiento del Cáncer, versión próstata (FACT-P), en pacientes mexicanos. *Cir Cir.* 2020;88(6):745-752. <https://doi.org/10.24875/CIRU.19001755>.
12. Chang C, Gardiner J, Houang R, et al. Comparing multiple statistical software for multiple-indicator, multiple-cause modeling: an application of gender disparity in adult cognitive functioning using MIDUS II dataset. *BMC Med Res Methodol.* 2020;20(1):275. <https://doi.org/10.1186/s12874-020-01150-4>.
13. Littlewood Zimmerman HF. *Mi primer modelamiento de ecuación estructural LISREL*. 3ª ed. México: DF; 2022 [consulta 2024 mar 01]. Disponible en: [https://www.researchgate.net/publication/364844144\\_Mi\\_primer\\_modelamiento\\_de\\_ecuacion\\_estructural\\_LISREL](https://www.researchgate.net/publication/364844144_Mi_primer_modelamiento_de_ecuacion_estructural_LISREL).
14. Zaky HHM, Armanious DM, Kalliny MA. Applying the structural equation model approach to study the simultaneous relationship between women's empowerment and mental disorder in Egypt. *BMC Womens Health.* 2024;24(1):26. <https://doi.org/10.1186/s12905-023-02863-6>.
15. Moral de la Rubia J, López Rosales F. Modelo recursivo de reacción violenta en parejas válido para ambos sexos. *Bol Psicol.* 2012;(105):61-74.