MIMIC Models: From Neuroscience to Health Sciences

María F. Grande Ratti[1], Romina Y. Pérez Manelli[1], Fernando R. Vázquez Peña[2] and Eulogio Cordón Pozo[3]

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

ABSTRACT

In medicine, situations often arise where regression (linear or logistic) is not feasible due to latent constructs or factors that include multiple observed variables as they attempt to reflect or estimate complex multidimensional phenomena. This limitation has led to the search for alternative statistical methodologies, such as a subtype of Structural Equation Models known as MIMIC models (Multiple Indicators, Multiple Causes), initially used in neuroscience. This article explains what these models are, their purpose, the software available for their application, and how to interpret the results. An example is provided based on a study on the quality of life of individuals who have undergone a transplant, highlighting graphical representation and the importance of parsimony in modeling. These models allow for the simultaneous inclusion of factors and individual variables, addressing complex causal relationships and making them valuable tools in medical research.

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

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

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.

Author for correspondence: maría.grande@hospitalitaliano.org.ar, Grande Ratti MF.

Received: 07/16/24 Accepted: 08/27/24 Online: 10/10/24

DOI: http://doi.org/10.51987/revhospitalbaires.v44i3.344

How to cite: Grande Ratti MF, Pérez Manelli RY, Vázquez Peña FR, Cordón Pozo E. MIMIC Models: From Neuroscience to Health Sciences. Rev. Hosp. Ital. B. Aires. 2024;44(4)

ISSN (online) 2314-3312

INTRODUCTION

Cognitive neuroscience has inspired a series of methodological advances in health by employing popular techniques to summarize behavioral data that include scores, where the dimensionality of items and the quality of information often remain unexplored in depth[1]. For example, the PHQ-9 instrument is widely used to diagnose and determine the severity of depression in primary care[2]. However, the total score results from summing the value of each response, and cutoff points exist to classify mild, moderate, moderately severe, and severe depression (with 5, 10, 15, and 20 points, respectively).

In this context, nine independent items are collected, whose content allows for the construction of a global continuous numerical variable, which is then interpreted and reported as a multiple and ordinal categorical variable. This process moves from granularity to constructs through transformations and the operationalization of variables.

Before moving forward, we will define "variables" and "factors" as they play key roles in structure and analysis. Factors refer to latent constructs that are not directly observable, but in practice, they are inferred from variables. They represent the underlying theoretical concepts that integrate them and, consequently, help reduce the amount of data to analyze. In contrast, variables are concrete and observable measurements that are directly recorded and used to measure those latent constructs.

Classically, they can be divided into two types: dependent (Indicators) because they indicate the state of the latent factor, and independent (Causes) because they influence the latent factor. For example, if the factor were "mental health," the indicators could include survey responses about anxiety and depression, while the causes could contain variables such as stress levels and economic status. Therefore, factors are influenced by causes and measured through indicators.

Additionally, we will review the classification of variables based on their nature (or type of data). They can be: A) Qualitative or Categorical, which describe qualities or characteristics, and either Nominal (without intrinsic order, e.g., male/female) or Ordinal (with a hierarchy, e.g., educational level); B) Quantitative or Numerical, which represent quantities and can be either Discrete (taking whole and finite values, e.g., number of children) or Continuous (taking any value within a range, e.g., age, weight, or blood glucose level).

In medicine, the use of dichotomous variables is very frequent, often with an overly simplistic approach (e.g., hospitalization yes/no). Dichotomous (or binary) variables can only take two distinct values (e.g., disease present or absent). Consequently, we often encounter situations where we would like to design a model that allows for the simultaneous inclusion of factors (composed of various variables) and individual variables (e.g., gender, age). In these cases, we might consider using tools such as regression (logistic or linear), but the presence of factors prevents us from doing so. Alternatively, we could explore developing a classical Structural Equation Model (SEM– a multivariate statistical technique used to test and estimate causal relationships based on data and qualitative assumptions about causality), but we cannot do it for the single variables either.

Linear or logistic models are not always appropriate in all scenarios due to certain inherent limitations of these techniques and the underlying assumptions. Although both methods are used for making predictions, they differ in the type of dependent variable they model. Linear regression is employed to model the relationship between a continuous dependent variable and one or more independent variables (which can be either continuous or categorical). Logistic regression is employed to model the probability of occurrence of a binary event based on one or more independent variables. However, both methods involve a dependent variable and one or more independent variables. They are all observed variables, and neither technique works with latent factors or variables. They do not allow: A) a variable to be both dependent and independent simultaneously within the same model, or B) one variable to influence another while simultaneously allowing the latter to affect the former (models referred to as "non-recursive"). Furthermore, traditional models do not allow working with individual (non-factorized) variables.

The alternative that nowadays allows addressing this complexity is using the MIMIC models (from *Multiple Indicators, Multiple Causes*). Although they have been successfully employed in mental health[3] and geriatric[4] research, they have not yet been adopted in many other areas where they have not yet been applied, despite being a well-known methodological approach for quite some time[5]. However, the possibility of using them with ordinal variables in the Lisrel program became widely known mainly through Jöreskog's conferences, held in Germany in 2001. Taking this event as a starting point, the evolution of asymptotically distribution-free estimators (from the English *ADF*, or *Asymptotically Distribution-Free*) has been crucial. These do not require the assumption of multivariate normality and, therefore, allow working with ordinal variables, which usually do not meet this assumption. The first was Weighted Least Squares (*WLS*) in Lisrel; later, Diagonally Weighted Least Squares (*DWLS*)[6] emerged, allowing for work with smaller sample sizes.

Currently, in Mplus and R Lavaan, we find WLSMV or Robust DWLS[7], which achieve acceptable performance under all conditions, even with small sample sizes, with a minimum of 100 observations[8].

What are MIMIC models, and what are they used for?

MIMIC models are SEMs that enable us to work with factors (latent variables) and individual variables (non-factorized), providing the possibility of performing multiple regressions with both factors and variables at the same time. This technique involves the use of latent variables that are predicted by observed variables[9].

The main idea behind MIMIC models is to incorporate observed variables –whether dependent or independent– into the model alongside latent variables. To understand this, we must first recall the definitions of latent and observed variables. Latent variables are those that are not directly observed but are assumed to influence observed variables, while observed variables are those that are directly measured. For example, one study mentions the construct of *dyslipidemia* as being made up of different serum cholesterol measurements[10] (Fig. 1).

Why is it important to work with factors?

When estimating constructs or complex indicators, relying on a single variable is often not enough. For example, can we study *metabolic syndrome*, *quality of life*, or *healthy habits* with just one variable? In clinical practice, we commonly approach these phenomena using a multidimensional framework, which involves measuring and considering various variables (factors) because they are complex concepts.

Sometimes, we also want to provide more stability to the model and, thus, prefer using factors. For instance, if we are using *total cholesterol level* as an independent variable, it would be more stable to use *a lipid profile*, including LDL and triglyceride measurements, if conceptually appropriate.

However, in other disciplines (such as psychology or business administration), it is more common to find studies that rely exclusively on factors and hence use classical SEM models. In medicine, however, we almost always work with variables that are difficult to factorize, at least in most clinical studies –such as age or gender– or dichotomous variables (e.g., deceased or not).

Illustration with a Real Example

A cross-sectional study included the self-reported perception of 361 individuals who had received a solid organ transplant, assessing quality of life using tools such as EuroQol (a health-related quality of life measurement instrument in which each person evaluates their health status) and SF-36v2 (another health-related quality of life measurement instrument, where individuals also assess their health status). Additionally, the study included non-factorized observable variables such as age, gender, educational level, number of transplants, and years since the last transplant.

The first instrument includes a Visual Analog Scale to assess health-related quality of life. This tool allows respondents to express their subjective perception of their health by placing a mark on a horizontal line ranging from 0 (representing the worst imaginable health) to 100 (representing the best imaginable health).

The second instrument consists of 36 questions grouped into eight health dimensions: general health, vitality, physical functioning, physical role, bodily pain, mental health, emotional role, and social functioning.

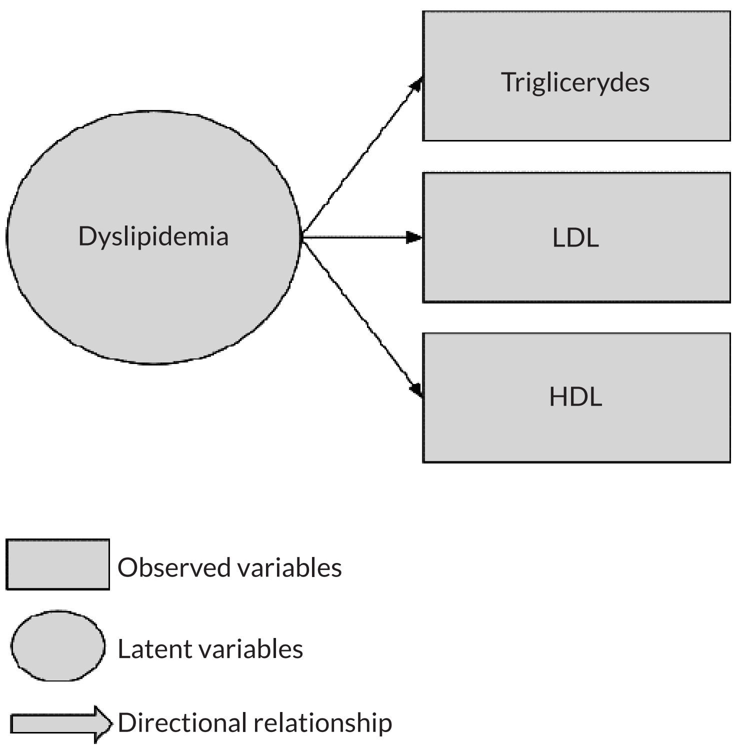

Figure 1. Latent construct (*dyslipidemia*) based on observed variables (cholesterol measurements). LDL: Low-Density Lipoprotein. HDL: High-Density Lipoprotein. Source: Own elaboration, adapted from the original[10].

The crucial first step is to design and outline a diagram, typically defining the model structure, including the connections and causal relationships between latent and observed variables. That requires graphically outlining the model to understand the relationship between a latent variable and its observed indicators, also considering the influence of external variables or covariates. In the graphical representation (Fig. 2), latent variables are typically depicted as ovals or circles (in blue), while observed variables (indicators) appear as rectangles (in green); arrows between them indicate causal relationships.

Parsimonious Model

This term refers to the preference for simpler and more concise models when explaining data without significantly sacrificing goodness of fit, aiming to avoid unnecessary complexity[11]. In other words, the goal is to find a model that balances fitting the data well while being as simple as possible.

Statistical Software

MIMIC models can be estimated using various statistical software packages, such as R Lavaan, Stata, MPlus, and SAS, where the results obtained are highly convergent[12]. The specification and straightforward testing of the MIMIC models can be divided into three successive steps.:

1) Draw the MIMIC model, which allows for a clear visualization of its structure and how latent variables and observed indicators are interrelated, facilitating the understanding and interpretation of the model.
2) Develop a parsimonious model that fits well for the sample as a whole, ensuring that the indices are significant and of sufficient magnitude.
3) Carefully interpret and generalize the results, considering the model's limitations and assumptions. That involves evaluating how well the data "fit" the proposed model based on information provided by the chi-square test, ad hoc indices, and the significance of estimated parameters and/or coefficients. We know that, in regression-based studies of this type, it is advisable to use standardized coefficients to avoid interpretation issues related to the units used for categorizing variables (Fig. 3). In this case, the standardized coefficients of MIMIC models are viewed as the beta estimators in linear regressions.

Advantages of SEM

Using MIMIC models provides additional advantages that are not widely known in the medical field, although they are frequently used in other disciplines:

- Variables can be both independent and dependent at the same time. That is, there are no exclusively independent or exclusively dependent variables. A variable or factor A can be independent in its relationship with factor B while simultaneously dependent on variable Z[13].
- Non-recursive models also exist. In these models, variable A influences variable B, and at the same time, variable B influences variable A. These models, for example, are used to study domestic violence. In contrast, the classic recursive model is one in which causal relationships between variables are

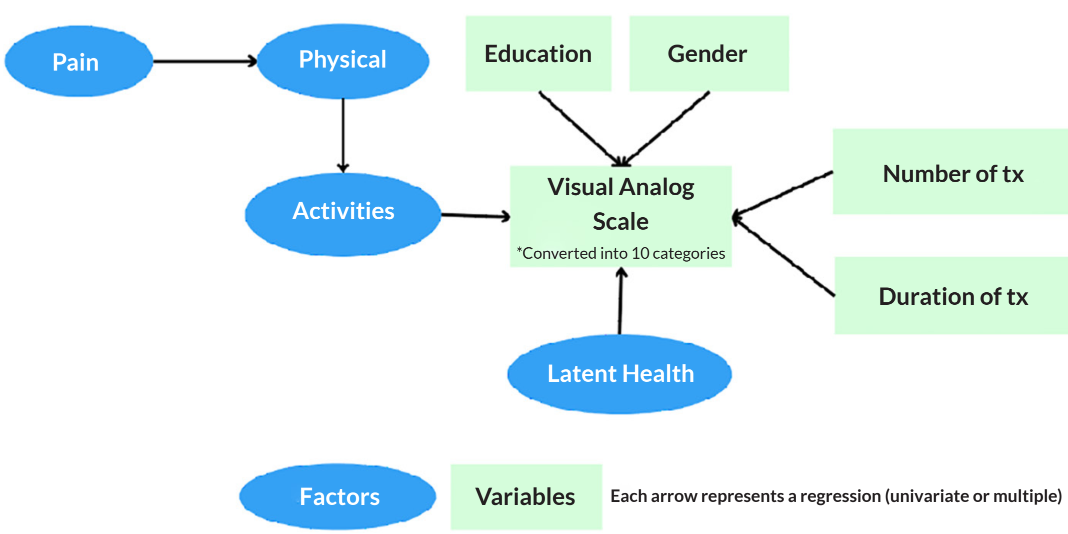

Figure 2. One of the graphical representations of a real project based on self-reported quality of life by individuals who have received a transplant (tx). Source: Own elaboration.

Regressions:

	Estimate	Std.Err	z-value	P(>|z|)	Std.lv	Std.all
Health						
How I feel	-0.099	0.083	-1.192	0.233	-0.071	-0.070
activities	0.157	0.078	2.015	0.044	0.122	0.121
latent health	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
activities -						
emotional	0.041	0.053	0.774	0.439	0.050	0.050
physical	0.649	0.054	12.070	0.000	0.744	0.744
emotional –						
pain	-0.698	0.048	-14.428	0.000	-0.627	-0.627
physical –						
pain	-0.990	0.042	-23.417	0.000	-0.944	-0.944

Figure 3. One of the results of the MIMIC model (in the real case) using R Lavaan. We are mainly interested in the values returned by the columns "Std. all" (standardized coefficients) and "P (>|z|)" (p-value). x11: formal studies of the patients, in years; x5: age of the patients, in years; x3: gender of the patients

unidirectional and acyclic, meaning there is no feedback between variables[14,15].

CONCLUSIONS

MIMIC models are useful analytical tools for understanding and modeling complex relationships between latent and observed variables, considering the possibility of multiple causal factors.

Funding: This project did not receive any funding.

Conflicts of Interest: The authors declare no conflicts of interest.

Author Contributions: Conceptualization, Data Curation, Formal Analysis, Reserch, Software: MFGR, RYPM, FRVP. Methodology, Project Administration: MFGR and FRVP. Supervision, Validation, Visualization, Writing – Original Draft, Writing – Review & Editing: MFGR, RYPM, FRVP, ECP.

REFERENCES

1. 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.
2. 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.
3. 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.
4. 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.
5. 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.
6. 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.
7. 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.
8. 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.

9. Schumacker RE, Lomax RG. A beginner's guide to structural equation modeling. 3rd ed. New York, NY: Routledge; 2012.
10. 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.
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.