

## Análisis multivariado

Diego Pérez de Arenaza 

Jefe de Sección de Imágenes Cardiovasculares del Servicio Cardiología, Hospital Italiano de Buenos Aires.  
Ciudad Autónoma de Buenos Aires, Argentina.

*Acta Gastroenterol Latinoam* 2022;52(2):115-119

Recivido: 03/05/2022 / Aceptado: 05/06/2022 / Publicado online: 27/06/2022 / <https://doi.org/10.52787/agl.v52i2.206>

Muchos eventos –médicos, políticos y sociales– poseen múltiples causas, muchas de las cuales se relacionan unas con otras.

El análisis multivariado es una herramienta estadística para determinar la contribución relativa de diferentes causas a un determinado evento o resultado. Los investigadores clínicos necesitan este tipo de análisis, dado que las enfermedades presentan múltiples causas y el pronóstico se determina, usualmente, por un gran número de factores.

Esta herramienta va a ser aplicada en dos escenarios clínicos mediante dos ejemplos.

### Ejemplo 1

En la enfermedad coronaria concurren varios factores, como el tabaquismo, la hipertensión arterial, la dislipemia, la diabetes y el antecedente coronario, que se asocian al infarto agudo de miocardio.

Hay que notar que estos factores *no causan* el infarto de miocardio, sino que *se asocian* a él.

La causalidad se basa en la plausibilidad biológica y en diseños rigurosos de estudio, como los estudios aleatorizados, que eliminan las fuentes potenciales de sesgo.

Sin embargo, la identificación de factores de riesgo asociados con el evento, a través de estudios observacionales, es particularmente importante, ya que no se puede aleatorizar a personas con diferentes condiciones que causan el infarto de miocardio.

A su vez, estas condiciones se asocian conjuntamente, ya que un paciente fumador puede ser hipertenso y presentar antecedentes coronarios.

En nuestra base de infarto agudo de miocardio (IAM), el objetivo es evaluar si la insuficiencia cardíaca (ICC) se asocia en forma independiente a la muerte en esta población de pacientes.

Lo que deseamos saber es si la ICC durante la internación contribuye en forma independiente a la mortalidad en el infarto agudo de miocardio.

La Tabla 1 compara las características de los pacientes que fallecieron vs. los que sobrevivieron en el post-IAM.

Podemos ver que hay muchas diferencias entre los grupos: los pacientes que fallecen son más añosos, presentan mayor porcentaje de ICC, episodios de taquicardia ventricular, bloqueo A-V y fibrilación ventricular.

Si observamos la Tabla 1, nos indica que los pacientes que padecen ICC post-IAM presentan una mayor mortalidad con respecto a los que no la presentan (51% vs. 28%).

Sin embargo, no nos responde la pregunta inicial acerca de la contribución independiente de la ICC, ya que solo se analiza la relación de la variable con el evento (análisis univariado) y no con los demás factores del estudio.

**Correspondencia:** Diego Pérez de Arenaza  
Correo electrónico: [diego.perezdearenaza@hospitalitaliano.org.ar](mailto:diego.perezdearenaza@hospitalitaliano.org.ar)

Por lo tanto, necesitamos el análisis multivariado para responder la pregunta en cuestión.

En la Tabla 2 se expresan los resultados del análisis multivariado en el que se *ajusta* el peso de esa variable con respecto a los demás factores.

Lo que puede analizarse de este cuadro es que los pacientes con ICC post-IAM presentan un riesgo elevado de muerte (1,7/1 en relación con los que no padecieron ICC) y esto es independiente de las demás variables, lo que significa que tiene peso propio.

Podría haberse realizado un análisis estratificado por múltiples variables.

Este tipo de análisis permite valorar el efecto de un factor de riesgo sobre el evento mientras la otra variable permanece constante.

**Tabla 1.** Características basales de los pacientes que sobrevivieron y de los que fallecieron en el IAM

Variable	Vivos (n=2705)	Muertos (n=445)
Edad ± (DS)	66 (13)	75 (11)
Sexo Mac (%)	64%	52%
Angina inestable previa (%)	13%	5%
Angina estable previa (%)	11,4%	3%
<b>Insuficiencia cardíaca (%)</b>	<b>28%</b>	<b>51,5%</b>
Diabetes	24%	21%
EPOC (%)	2,8%	3,1%
Hipertensión (%)	39%	22,4%
Taquicardia (%)	7,1%	13%
Bloque A-V (%)	8%	12,7%
Fibrilación ventricular (%)	2,4%	10,8%
Revascularización (%)	37,6%	11%

**Tabla 2.** Análisis multivariado

Muerte	Odds ratio	p	[95% Conf. interval.]	
Inestable	0,6	0,049	0,3	0,9
Estable	0,5	0,060	0,2	1,0
<b>ICC</b>	<b>1,7</b>	<b>0,000</b>	<b>1,3</b>	<b>2,1</b>
Diabetes	0,8	0,180	0,6	1,1
EPOC	0,9	0,772	0,4	1,7
HTA	0,5	0,000	0,3	0,6
TV	1,7	0,003	1,2	2,5
AVblock	1,5	0,021	1,0	2,1
FibV	6,4	0,000	4,0	10,3
Cath	0,3	0,000	0,2	0,5
Revasc.	0,6	0,049	0,4	0,9
Sexo	0,8	0,142	0,6	1,0
Edad	2,1	0,000	1,6	2,8

AVBlock= bloqueo AV, Cath= cateterismo coronario, EPOC= enfermedad obstructiva crónica, FibV= fibrilación ventricular, HTA= hipertensión, TV= taquicardia ventricular, ICC= insuficiencia cardíaca, Revasc.= revascularización coronaria.

Consideremos la relación de la edad con la ICC en el impacto de la mortalidad post-IAM como se muestra en la Tabla 3.

**Tabla 3.** Análisis bivariado: edad e ICC en relación al evento

Edad	OR	95% Intervalo de confianza	M-H Weight
<= 75 años	2,7	1,9 3,7	19,4
> 75 año	1,6	1,2 2,2	35,6
Crude	2,6	2,1 3,2	
M-H combined	2,0	1,6 2,5	

La presencia de ICC post-IAM incrementa el riesgo de muerte independientemente de la edad, ya que en el estrato de menor o igual 75 años, el riesgo es de 2,7 /1, y en los mayores de 75 años el riesgo es de 1,66 /1. A pesar de la edad, el riesgo persiste elevado.

Este tipo de análisis estratificado tiene sus limitaciones: debiéramos estratificar por cada una de las variables presentes en el estudio; con cada estratificación que agregamos, incrementamos el número de subgrupos para evaluar individualmente la relación de ICC y muerte en el IAM.

Así, en algunos subgrupos, habrá un número de pacientes insuficiente, aun comenzando con tamaños muestrales grandes.

Por lo tanto, solo evaluamos la ICC y la ajustamos a la edad, donde el riesgo de ICC es independiente de la edad, pero no la ajustamos a las demás variables que son importantes en la evolución del IAM.

El análisis multivariado resuelve esta limitación, ya que permite evaluar simultáneamente el impacto de múltiples variables en el resultado.

A continuación, enunciaremos los usos más comunes del análisis multivariado:

1. Identificar factores pronósticos, ajustando por potenciales confundidores: si bien los análisis multivariados son herramientas para el ajuste por potenciales confundidores, no hay que asumir que, si se realiza este modelo, se elimina por completo el sesgo de ese confundidor. Ningún ajuste es perfecto, dado que estos modelos presentan errores y, a su vez, pueden omitirse variables importantes, incorrectamente especificadas o interacciones entre las variables que no tomamos en cuenta.

2. Ajustar por las diferencias en las características basales: cuando la aleatorización es imposible, el uso del análisis multivariado aproxima estadísticamente una comparación de grupos “similares”.

3. Determinar modelos pronósticos: los modelos pronósticos proveen una válida estimación del riesgo solo en pacientes con características similares a la población estudiada.

4. Determinar modelos diagnósticos: los modelos multivariados pueden identificar la mejor combinación de información diagnóstica para una persona con una enfermedad en particular.

## Ejemplo 2

En el siguiente ejemplo se analiza si el reflujo gastroesofágico (RGE) es un factor de riesgo para el adenocarcinoma de esófago sobre la base del siguiente artículo científico: N Engl J Med. 1999;340(11):825-31.

## Diseño

*Caso control*, apareados por edad y sexo en estratos por diez años. Es un diseño cuya dirección va del caso (cáncer) al predictor (RGE), ya que los autores recolectaron a pacientes con cáncer e identificaron cuántos padecieron síntomas típicos de RGE durante los cinco años previos.

## Metodología

- Identificación minuciosa y uniforme de los casos representados por pacientes con cáncer esofágico (adenocarcinoma de esófago, adenocarcinoma del cardias y carcinoma escamoso de esófago).

- **Controles poblacionales:** apareados por edad y sexo, seleccionados en forma aleatoria de un registro representativo de la población de Suecia. Además, se tomaron como controles pacientes con carcinoma epidermoide de esófago.

- **Casos:** fueron seleccionados en centros de cirugía y de un registro nacional de cáncer de Suecia, para tratar de reclutar la mayor parte de los pacientes con tumores de esófago (adenocarcinoma).

## Análisis

Univariado y multivariado por regresión logística (expresado a través del “odds ratio”, que nos indica la relación de riesgo).

Se ajustó en el análisis por once potenciales confundidores (que los autores estimaron relevantes para identificar la independencia de la variable en estudio [RGE] con relación al resultado [cáncer de esófago]).

## Resultados

La Tabla 1 expresa las características de la población global en estudio (edad, sexo, consumo de tabaco, de alcohol, nivel de educación) que, según los autores, adquieren relevancia en relación con esta patología.

La Tabla 4 expresa el riesgo de los síntomas de RGE en una semana o por la noche en una semana (esta tabla indica el predictor que estamos evaluando en relación con el evento).

El análisis es *univariado*, ya que es una sola variable (reflujo) en relación con el evento (cáncer).

Una primera mirada a la Tabla 2 la focalizamos en los controles (entre un 84-92% no padecen síntomas de reflujo), que nos muestra una elección adecuada de estos.

En una segunda interpretación en la que en los diferentes subgrupos de cáncer se expresa una relación entre no padecer síntomas (Odds ratio = 1 [grupo comparativo]) y presentar estos síntomas (cuántas veces se eleva el riesgo de cáncer al presentarlos vs. no padecerlos).

Ejemplo Tabla 4 (véase en el texto)

	Controles (n= 820)	Adenocarcinoma de esófago (n= 189)	
Síntomas de reflujo			
No, n (%)	685 (84)	76 (40)	761
Sí, n (%)	135 (16)	113 (60)	248
	820	189	1009

Odds ratio (OR):

$$\frac{\text{Controles sin síntomas} \times \text{Casos con síntomas}}{\text{Controles con síntomas} \times \text{Casos sin síntomas}} = \frac{685 \times 113}{135 \times 76} = 7,7$$

Si vamos al texto, se presenta como OR:1 (en controles y casos que no tienen síntomas) y un OR de 7,7 (5,3-11,4) en aquellos que los presentan.

Los pacientes que presentan síntomas de reflujo tienen siete veces más riesgo de cáncer de esófago que aquellos que no presentan síntomas de RGE.

No obstante, este análisis es solo de la variable en cuestión versus el evento; sin embargo, hay otras variables confundidoras de esta relación que los autores habían preestablecido.

Por consiguiente, realizaron un análisis multivariado (Tabla 3) en el que el OR no ajustado (univariado) del síntoma de RGE para el riesgo de adenocarcinoma esofágico es de 7,5 (el intervalo de confianza [IC]: 5,4-10,6); el de adenocarcinoma del cardias gástrico es de 2,0 (IC: 1,5-2,8), y el de carcinoma escamoso de esófago, de 0,9 (IC: 0,6-1,4).

Presentar síntomas típicos de RGE representa un riesgo elevado de padecer adenocarcinoma de esófago y de cardias gástrico (IC no pasa por el valor 1), pero no de carcinoma escamoso (el IC pasa por el valor de nulidad).

En esta tabla 5, por debajo del valor no ajustado se describen otras variables potencialmente confundidoras de esta relación (RGE y cáncer).

**Tabla 5.** Valor no ajustado y ajustado por potenciales confundidores de la relación del reflujo RGE y el cáncer

VARIABLES confundidoras	Adenocarcinoma de esófago	Adenocarcinoma del cardias gástrico	Carcinoma esofágico de células escamosas
Odds ratio (95% intervalo de confianza)			
No ajustado	7,5 (5,4 – 10,6)	2,0 (1,5 – 2,8)	0,9 (0,6 – 1,4)
Edad	7,6 (5,3 – 10,7)	2,0 (1,4 – 2,8)	0,9 (0,6 – 1,5)
Sexo	7,5 (5,3 – 10,6)	2,0 (1,4 – 2,8)	1,0 (0,6 – 1,6)
Índice de masa corporal	7,1 (4,9 – 10,2)	1,9 (1,4 – 2,7)	1,0 (0,6 – 1,6)
Tabaco	7,1 (4,9 – 10,2)	2,0 (1,4 – 2,7)	1,1 (0,7 – 1,8)
Alcohol	7,6 (5,2 – 11,0)	2,0 (1,4 – 2,8)	1,2 (0,7 – 2,1)
Nivel de educación	7,6 (5,2 – 11,0)	2,0 (1,4 – 2,8)	1,2 (0,7 – 2,0)
Consumo energético	7,6 (5,2 – 11,1)	2,0 (1,4 – 2,8)	1,2 (0,7 – 2,0)
Consumo de frutas y vegetales	7,6 (5,2 – 11,1)	2,0 (1,4 – 2,8)	1,2 (0,7 – 2,0)
Trabajar en una postura encorvada	7,6 (5,2 – 11,1)	2,0 (1,4 – 2,8)	1,2 (0,7 – 2,0)
Actividad física en el trabajo	7,8 (5,3 – 11,4)	2,0 (1,4 – 2,9)	1,1 (0,7 – 1,9)
Actividad física en el tiempo libre	7,7 (5,3 – 11,4)	2,0 (1,4 – 2,9)	1,2 (0,7 – 2,0)

Estas variables potencialmente confundidoras (once en total) se enuncian con su respectivo OR, que no es el de la variable en cuestión, sino cómo hacen variar el OR no ajustado del reflujo GE en relación con las neoplasias; así, la edad no hace variar el OR no ajustado para el carcinoma de esófago (7,6 con la edad vs. 7,5 no ajustado), cuando incorporamos el sexo no varía el OR y así sucesivamente hasta llegar a la última variable (actividad física durante el esparcimiento), que no hace variar el OR e incorpora todas las demás.

Es decir, no se confunde la relación del RGE con el adenocarcinoma de esófago, tomando estas once variables.

En relación con el adenocarcinoma de cardias, la relación tampoco se altera, estimando el riesgo en dos veces más que en aquellos que no presentaron reflujo de forma independiente (no está confundida la relación por las demás variables).

En relación con el carcinoma escamoso, en el OR no ajustado, no hay relación con el RGE y esta no asociación se mantiene aun con los potenciales confundidores.

En resumen, el estudio demostró el riesgo mayor de adenocarcinoma de esófago y de cardias con el RGE. La relación de riesgo fue mayor para el adenocarcinoma de esófago.

No está confundida esta relación por las otras variables (valor independiente).

El RGE no es un factor de riesgo del carcinoma de células escamosas.

Las fortalezas del estudio son:

1. Estudio caso control con una adecuada selección del caso y un apareamiento de los controles por edad y sexo.

2. Se realizó un análisis por potenciales confundidores biológicamente vinculados a la patología en cuestión.

Las debilidades del estudio son:

1. En un estudio caso control hay que tener en cuenta los sesgos de recolección de datos (fueron adecuadamente extraídos de bases validadas; siempre está la posibilidad de tener una información no del todo completa de los casos).

2. Se establecieron 11 potenciales confundidores, pero uno puede inferir que puede haber otros no incluidos que hubiera que incorporar (en esto el estudio aleatorizado ajusta y equipara por potenciales confundidores conocidos y desconocidos a diferencia de los observacionales).

Este trabajo muestra en forma definitiva la relación causal entre el adenocarcinoma de esófago y la enfermedad por reflujo.

**Propiedad intelectual.** El autor declara que los datos y las tablas presentes en el manuscrito son originales y fueron realizados en su institución perteneciente.

**Financiamiento.** El autor declara que no hubo fuentes de financiación externas.

**Conflicto de interés.** El autor declara no tener conflictos de interés en relación con este artículo.

#### Aviso de derechos de autor



© 2022 Acta Gastroenterológica Latinoamericana. Este es un artículo de acceso abierto publicado bajo los términos de la Licencia Creative Commons Attribution (CC BY-NC-SA 4.0), la cual permite el uso, la distribución y la reproducción de forma no comercial, siempre que se cite al autor y la fuente original.

**Cite este artículo como:** Pérez de Arenaza D. Análisis multivariado. Acta Gastroenterol Latinoam. 2022; 52 (2): 115-119. <https://doi.org/10.52787/agl.v52i2.206>

#### Bibliografía

1. Mitchell H. Katz. Multivariable Analysis: A Practical Guide for Clinicians and Public Health Researchers. CAMBRIDGE UNIVERSITY PRESS. ISBN: 9780521760980.
2. Charles Hennekens and Julie E. Buring, Epidemiology in Medicine. Lippincott Williams and Wilkins. ISBN: 9780316356367.
3. Lagergren J, Bergström R, Lindgren A, Nyrén O. Symptomatic gastroesophageal reflux as a risk factor for esophageal adenocarcinoma. N Engl J Med. 1999 Mar 18;340(11):825-31