

## Multivariate Analysis

Diego Pérez de Arenaza 

Head of Cardiovascular Imaging Section, Cardiology Department, Hospital Italiano de Buenos Aires.  
Ciudad Autónoma de Buenos Aires, Argentina.

*Acta Gastroenterol Latinoam* 2022;52(2):120-124

Received: 03/05/2021 / Accepted: 05/06/2021 / Published online: 27/06/2022 / <https://doi.org/10.52787/agl.v52i2.206>

Many medical, political and social events have multiple causes, many of them related to each other.

Multivariate analysis is a statistical tool to determine the relative contribution of different causes to a given event or outcome. Clinical researchers need this type of analysis, since diseases have multiple causes and prognosis is usually determined by a large number of factors.

This type of analysis will be applied in two clinical scenarios by means of two examples.

### Example 1

Coronary heart disease is associated with several factors, such as smoking, arterial hypertension, dyslipidemia, diabetes and coronary history, which are associated with acute myocardial infarction.

It should be noted that these factors *do not cause* myocardial infarction, but *are associated* with it.

Causality is based on biological plausibility and rigorous study designs, such as randomized trials which eliminate potential sources of bias.

However, the identification of risk factors associated with the event through observational studies, is particularly important, since it is impossible to randomize people with different conditions that cause myocardial infarction.

In turn, these conditions are associated together, since a smoker patient can be hypertensive and have a coronary history.

In our acute myocardial infarction (AMI) database, the objective is to assess whether heart failure (CHF) is independently associated with death in this patient population.

What we want to know is whether CHF during hospitalization independently contributes to mortality in acute myocardial infarction.

Table 1 compares the characteristics of the patients who died vs. those who survived post-AMI.

We can see that there are many differences between the groups: patients who died were older, had a higher percentage of CHF, episodes of ventricular tachycardia, AV block and ventricular fibrillation.

Table 1 shows that patients with post-AMI CHF have higher mortality compared to those who do not (51% vs. 28%).

However, it does not answer the initial question about the independent contribution of the CHF, since it only analyzes the relationship of the variable with the event (univariate analysis) and not with the other factors in the study.

---

**Correspondence:** Diego Pérez de Arenaza  
Email: [diego.perezdearenaza@hospitalitaliano.org.ar](mailto:diego.perezdearenaza@hospitalitaliano.org.ar)

Therefore, we need multivariate analysis to answer the question at hand.

Table 2 presents the results of the multivariate analysis in which the weight of this variable is adjusted with the other factors.

What can be analyzed from this table is that patients with post-AMI CHF have an elevated risk of death (1.7/1 compared to those who did not suffer CHF) and this is independent of the other variables, which means it has its own weight.

A *multivariable stratified analysis* could have been performed.

This type of analysis allows the effect of one risk factor on the event to be assessed while the other variable remains constant.

**Table 1.** Baseline Characteristics of Patients with AMI

Variable	Survivors (n=2705)	Non Survivors (n=445)
Age ± (DS)	66 (13)	75 (11)
Male Sex (%)	64%	52%
Previous Unstable Angina (%)	13%	5%
Previous Stable Angina (%)	11.4%	3%
<b>Congestive Heart Failure (%)</b>	<b>28%</b>	<b>51.5%</b>
Diabetes	24%	21%
COPD (%)	2.8%	3.1%
Hipertension (%)	39%	22.4%
Ventricular Tachycardia (%)	7.1%	13%
AV Block (%)	8%	12.7%
Ventricular Fibrillation (%)	2.4%	10.8%
Revascularization (%)	37.6%	11%

**Table 2.** Multivariate Analysis

Died	Odds ratio	p	[95% Conf. interval.]	
Unstable	0.6	0.049	0.3	0.9
Stable	0.5	0.060	0.2	1.0
<b>CHF</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
COPD	0.9	0.772	0.4	1.7
HTN	0.5	0.000	0.3	0.6
VT	1.7	0.003	1.2	2.5
AVblock	1.5	0.021	1.0	2.1
VFib	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
Gender	0.8	0.142	0.6	1.0
Age	2.1	0.000	1.6	2.8

**AV Block:** atrioventricular block, **Cath:** coronary catheterization, **COPD:** chronic obstructive pulmonary disease, **FibV:** ventricular fibrillation, **HTN:** hypertension, **VT:** ventricular tachycardia, **CHF:** congestive heart failure, **Revasc:** coronary revascularization.

Let us consider the relationship of age with CHF in the impact of post-AMI mortality as shown in Table 3.

**Table 3.** Bivariate analysis age and ICC in relation to the event

Age	OR	95% Confidence interval	M-H Weight
<= 75 years	2.7	1.9 3.7	19.4
> 75 years	1.6	1.2 2.2	35.6
Crude	2.6	2.1 3.2	
M-H combined	2.0	1.6 2.5	

The presence of post-AMI CHF increases the risk of death regardless of age, since in the stratum of less than or equal to 75 years, the risk is 2.7/1, and 1.66/1 in those older than 75 years. Despite age, the risk remains elevated.

This type of stratified analysis has its limitations: we should stratify by each of the variables present in the study, with each stratification we add, we increase the number of subgroups to individually assess the relationship of CHF and death in AMI.

Thus, in some subgroups, there will be an insufficient number of patients even starting with large sample sizes.

Therefore, we only assessed CHF and adjusted for age, where the risk of CHF is of age, but we did not adjust for the other variables that are important in the evolution of AMI.

Multivariate analysis resolves this limitation, as it allows the impact of multiple variables on the outcome to be assessed simultaneously.

The following are the most common uses of multivariate analysis:

1. Identify prognostic factors, adjusting for potential confounders: Although multivariate analyses are tools for adjusting for potential confounders, it should not be assumed that, if this model is performed, the bias of that cofounder is eliminated. No adjustment is perfect, since these models have errors and, in turn, may omit important and incorrectly specified variables or interactions between them that we do not consider.

2. Adjusting for differences in baseline characteristics: When randomization is impossible, the use of multivariate analysis statistically approximates a comparison of "similar" groups.

3. Determine prognostic models: Prognostic models provide a valid estimate of risk only in patients with similar characteristics to the population studied.

4. Determine diagnostic models: Multivariate models can identify the best combination of diagnostic information for a person with a particular disease.

## Example 2

The following example discusses whether gastroesophageal reflux (GER) is a risk factor for adenocarcinoma of the esophagus based on the following article: *N Engl J Med.* 1999;340(11):825-31.

## Design

Case control matched by age and sex in strata for 10 years: it is a design whose direction goes from the case

(cancer) to the predictor (GER), as the authors collected patients with cancer and identified how many had typical GER symptoms during the previous five years.

## Methodology

- Thorough and uniform identification of cases represented by patients with esophageal cancer (adenocarcinoma of the esophagus, adenocarcinoma of the cardia and squamous cell carcinoma of the esophagus).

- Controls matched by age and sex, randomly selected from a population registry representative of the Swedish population. In addition, patients with squamous cell carcinoma of the esophagus were taken as controls.

- Cases: they were selected from surgery centers and from a national cancer registry in Sweden, in an attempt to recruit the the majority of patients with esophageal tumors (adenocarcinoma).

## Analysis

Univariate and multivariate by logistic regression (expressed through the "odds ratio" which indicates the risk ratio).

The analysis was adjusted for eleven potential confounders (which the authors considered relevant to identify the independence of the variable under study [GER] in relation to the outcome [esophageal cancer]).

## Results

Table 1 shows the characteristics of the overall study population (age, sex, smoking, alcohol consumption and level of education) which, according to the authors, are relevant in relation to this pathology.

Table 4 expresses the risk of GER symptoms in one week or at night in one week (this table indicates the predictor that we are evaluating in relation to the event).

The analysis is *univariate*, since it is a single variable (reflux) in relation to the event (cancer).

A first look at Table 2 focuses on the controls (84%-92% do not have reflux symptoms), which shows an appropriate choice of controls.

In a second interpretation, a relationship is expressed, in the different cancer subgroups, between not having symptoms (Odds ratio=1 [comparative group]) and presenting these symptoms (how many times the risk of cancer is increased by presenting the symptoms vs. not having them).

If we look at the text, it is presented as OR:1 (in controls and cases that do not have symptoms) and an OR of 7.7 (5.3-11.4) in those who do.

Example table 4 (see in text)

	Controls (n=820)	Carcinoma of the esophagus (n=189)	
Symptoms of Reflux			
No, n (%)	685 (84)	76 (40)	761
Yes, n (%)	135 (16)	113 (60)	248
	820	189	1009

Odds ratio (OR):

Controls without symptoms

$$\frac{\text{x Cases with symptoms}}{\text{Controls with symptoms}} = \frac{685 \times 113}{135 \times 76} = 7.7$$

Controls with symptoms

x Cases without symptoms

Patients with reflux symptoms have a seven times higher risk of esophageal cancer than those without GER symptoms.

However, this analysis is only of the variable in question versus the event. There are other confounding variables in this relationship that the authors had pre-established.

Therefore, they performed a multivariate analysis (Table 3) where the unadjusted (univariate) OR of GER symptom for risk of esophageal adenocarcinoma is 7.5 (confidence interval [CI]: 5.4-10.6); that of adenocarcinoma of the gastric cardia is 2.0 (CI: 1.5-2.8); and that of squamous cell carcinoma of the esophagus is 0.9 (CI: 0.6-1.4).

Presenting symptoms typical of GER represents a high risk of esophageal and gastric cardia adenocarcinoma (CI does not go through the value 1), but not of squamous cell carcinoma (the CI goes through the null value).

In this table 5, below the unadjusted value, other potentially confounding variables of this relationship (GER and cancer) are described.

These potentially confounding variables (eleven in total) are stated with their respective OR, which is not that of the variable in question, but how they make the

**Table 5.** Unadjusted and adjusted value for potential confounders of the relationship of GE reflux and cancer

Confounding Variables	Esophageal Adenocarcinoma	Gastric Cardia Adenocarcinoma	Esophageal Squamous Cell Carcinoma
Odds ratio (95% intervalo de confianza)			
Unadjusted	7.5 (5.4 – 10.6)	2.0 (1.5 – 2.8)	0.9 (0.6 – 1.4)
Age	7.6 (5.3 – 10.7)	2.0 (1.4 – 2.8)	0.9 (0.6 – 1.5)
Sex	7.5 (5.3 – 10.6)	2.0 (1.4 – 2.8)	1.0 (0.6 – 1.6)
Body-mass index	7.1 (4.9 – 10.2)	1.9 (1.4 – 2.7)	1.0 (0.6 – 1.6)
Tobacco smoking	7.1 (4.9 – 10.2)	2.0 (1.4 – 2.7)	1.1 (0.7 – 1.8)
Alcohol use	7.6 (5.2 – 11.0)	2.0 (1.4 – 2.8)	1.2 (0.7 – 2.1)
Educational level	7.6 (5.2 – 11.0)	2.0 (1.4 – 2.8)	1.2 (0.7 – 2.0)
Energy intake	7.6 (5.2 – 11.1)	2.0 (1.4 – 2.8)	1.2 (0.7 – 2.0)
Intake of fruit and vegetables	7.6 (5.2 – 11.1)	2.0 (1.4 – 2.8)	1.2 (0.7 – 2.0)
Work in a stooped posture	7.6 (5.2 – 11.1)	2.0 (1.4 – 2.8)	1.2 (0.7 – 2.0)
Physical activity at work	7.8 (5.3 – 11.4)	2.0 (1.4 – 2.9)	1.1 (0.7 – 1.9)
Physical activity during leisure time	7.7 (5.3 – 11.4)	2.0 (1.4 – 2.9)	1.2 (0.7 – 2.0)

unadjusted OR of GER vary in relation to neoplasms; thus, age does not change the unadjusted OR for esophageal carcinoma (7.6 with age vs. 7.5 unadjusted), when we incorporate sex the OR does not vary and so on until we reach the last variable (physical activity during recreation), which does not vary the OR and incorporates all the others.

In other words, the relationship between GER and esophageal adenocarcinoma is not confused, taking these eleven variables.

In relation to adenocarcinoma of the cardia, the relationship is not altered either, estimating the risk to be twice as high as in those who did not present reflux independently (the relationship is not confounded by the other variables).

In relation to squamous cell carcinoma, in the unadjusted OR, there is no relationship with GER and this no-association is maintained even with the potential confounders.

*In summary*, the study demonstrated the increased risk of esophageal adenocarcinoma and cardia with GER. The risk ratio was higher for adenocarcinoma of the esophagus.

This relationship is not confounded by the other variables (independent value).

GER is not a risk factor for squamous cell carcinoma. The strengths of the study are:

1. Case-control study with adequate case selection and matching of controls by age and sex.
2. An analysis by potential confounders biologically linked to the pathology in question was performed.

Weaknesses of the study are:

1. In a case-control study, data collection biases must be taken into account (the data were adequately extracted from validated databases; there is always the possibility of having incomplete information on the cases).
2. Eleven potential confounders were established, but one can infer that there may be others not included that should be incorporated (in this the randomized study adjusts and equates for known and unknown potential confounders, unlike observational studies).

This study definitively shows the causal relationship between adenocarcinoma of the esophagus and reflux disease.

**Intellectual Property.** *The author declares that the data and tables that appear in this manuscript are original and were made in his belonging institution.*

**Funding.** *The author states that there were no external funding sources.*

**Conflict of interest.** *The author declares that he has no conflicts of interest in relation to this article.*

### Copyright



© 2021 Acta Gastroenterológica latinoamericana. This is an open-access article released under the terms of the Creative Commons Attribution (CC BY-NC-SA 4.0) license, which allows non-commercial use, distribution, and reproduction, provided the original author and source are acknowledged.

**Cite this article as:** Pérez de Arenaza D. Multivariate Analysis. *Acta Gastroenterol Latinoam.* 2022;52(2):120-124. <https://doi.org/10.52787/agl.v52i2.206>

### References

1. Katz M. H. *Multivariable Analysis: A Practical Guide for Clinicians and Public Health Researchers.* CAMBRIDGE UNIVERSITY PRESS. ISBN: 9780521760980.
2. Hennekens C., Buring J. E. *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