CARDIOVASCULAR AND METABOLIC SCIENCE www.ancam.org.mx

EPIDEMIOLOGY CORNER

doi: 10.35366/122137

Vol. 36 No. 4
October-December 2025

Uncertainty and probability are the basis of clinical epidemiology

La incertidumbre y la probabilidad son la base de la epidemiología clínica

José Luis Moragrega*

ABSTRACT

In science, including biology, complexity is the hallmark of the epoch. Statistics is the science of probability, and it is widely used since there is no certainty about the significance of the findings in an investigation, nor in a patient's data in the diagnosis or prognosis that we construct in clinical settings. The reductionism that medicine has employed to understand complex situations and construct the so-called «operative definition» to simplify complex problems must be revised to achieve a deeper understanding of the interactions among multiple factors influencing human bodily behavior.

RESUMEN

En la ciencia, incluyendo la biología, la complejidad es el sello distintivo de la época. La estadística es la ciencia de la probabilidad y es muy utilizada, ya que no hay certeza sobre la importancia de los hallazgos en la investigación ni en los datos de los pacientes en el diagnóstico o pronóstico que construimos en los entornos clínicos. El reduccionismo que la medicina ha elaborado para comprender muchas situaciones complejas y construir la llamada «definición operativa» para simplificar un problema que, de otro modo sería complejo, tiene que cambiarse por la comprensión profunda de las interacciones de muchos factores que influyen en el comportamiento del cuerpo humano.

Keywords: complexity, Bayes theorem, probability, science.

Palabras clave: *complejidad, teorema de Bayes, probabilidad, ciencia.*

INTRODUCTION

We know that statistics is the science of probability, and in relation to statistics and medicine, Sir William Osler wrote that, «Medicine is a science of uncertainty and an art of probability».[1]

In 2021, I published an article on how medicine is both a science and an art.[2] Now, as we define the role of statistics, we unveil another association between the two disciplines. In an era marked by complexity, as the late German physicist Hans-Peter Dürr[i] suggested, we must abandon reductionism and use statistics to comprehend natural phenomena. Albert Einstein held fast to his position on determinism, but Niels Bohr won out; Bohr's work confirmed Werner Heisenberg's uncertainty principle as the basis of quantum mechanics. You can read the beautiful book by Ian Stewart entitled «Does God Play Dice?»[3] and enjoy this beautiful dissertation.

Of course, in some moments it will rain if you are in the midst of a hurricane, and of course, you will die if you have rabies. However, in almost all other circumstances, events will occur with a certain probability. Medicine and meteorology are two fields where uncertainty is the norm. The simplest probabilistic statement will tell you the percentage probability of rain in the next 24 hours or the likelihood that pneumonia will resolve with a specific treatment. If it does not rain or the infection persists, this should not be surprising, given the probability of this occurrences.

* MD, FACC. Private practice of cardiology.
ORCID: 0009-0000-0084-4711

Received: 08/11/2025
Accepted: 11/06/2025

How to cite: Moragrega JL. Uncertainty and probability are the basis of clinical epidemiology. Cardiovasc Metab Sci. 2025; 36 (4): 228-232. https://dx.doi.org/10.35366/122137

[i] https://frasesdedios.blogspot.com/2016/11/la-realidad-ya-no-permite-reduccionismo.html

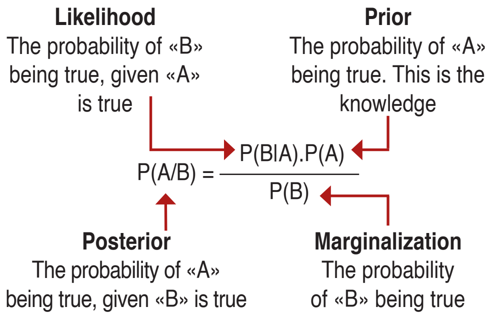

Figure 1: Conditional probability or Bayes' formula.

Table 1: Errors table. We can see that the concept is the same as accepting or rejecting the null hypothesis in scientific investigation.

	Reality	
Test	Sick	Healthy
	Accept alternative hypothesis H_1	Accept null hypothesis H_o
Sick Accept alternative hypothesis H_1	Correct ✓	Incorrect **Type I error** ✗
Healthy Accept null hypothesis H_o	Incorrect **Type I error** ✗	Correct ✓

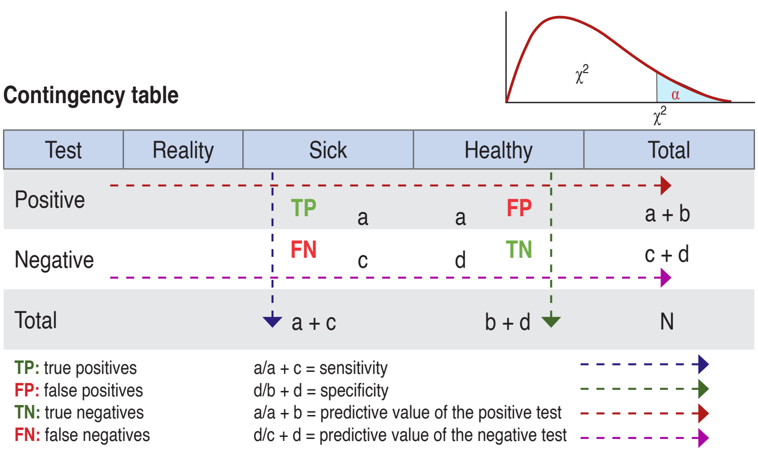

Figure 2: We can visualize the sensitivity (a/[a + c]) and specificity (d/[b + d]), along with the false positives (FP) and false negatives (FN), as shown by the vertical arrows. However, to determine the predictive value of a single patient, we calculate the proportion of positive (a/a + b) or negative (d/c + d) test results, as indicated by the horizontal arrows.

In biostatistics, we express the probability that the calculated sample values accurately reflect the parameters of the universe from which the sample is drawn. The most commonly used calculations are presented as confidence intervals or «p» values, where the probability can range from 0 to 1.

Probability values can be expressed as fractions, where the numerator represents the number of cases of a disease and the denominator represents the exposed population. For instance, the number of diabetics among obese males is 10/100. Furthermore, this can be express as 10%. This is the simplest way to measure uncertainty. However, in biology, the effect of the independent variable (obesity) on the dependent variable (diabetes) is always influenced by other factors. In addition to body weight, we can calculate the impact of age, gender, physical activity, genetics, and many other factors. As you can see, we are entering a complex realm, one that is no longer bound by the reductionism that has plagued medicine thus far. Conditional probability is a statistical measure that indicates the probability of an event «A» occurring if another event «B» has happened. That is the conditional probability P(A|B) *(Figure 1)*. In clinical medicine, we encounter multifactorial events every day, and it remains a mystery how medical professionals evaluate the information and make decisions. For example, we can apply an algorithm to a patient with chest pain to categorize them by gender, age group, and type of pain. This helps us determine the probability that the pain is ischemic.

A situation where the concept of conditional probability is beneficial is in the evaluation of laboratory and imaging tests. Our example will help illustrate that the values of sensitivity and specificity are not sufficient to assess the usefulness of a test. We must calculate the predictive value of a positive or negative test and then estimate the prevalence of the disease in the group to which an individual patient belongs (the prior probability). We then use Bayes' theorem, which involves the proportion of probabilities (also known as the likelihood ratio), to calculate the posterior probability.

	Myocardial infarction		
Troponin	Sick	Healthy	
+	TP A 95 (0.95) SENS	B FP 20 (0.2)	C 115
–	D FN 5 (0.05)	SPEC E 80 (0.8) TN	F 85
	G 100	H 100	200

Figure 3: Shows the published sensitivity (SENS) (95%) and specificity (SPEC) (80%) of the troponin test for myocardial infarction in patients with acute chest pain.[8]

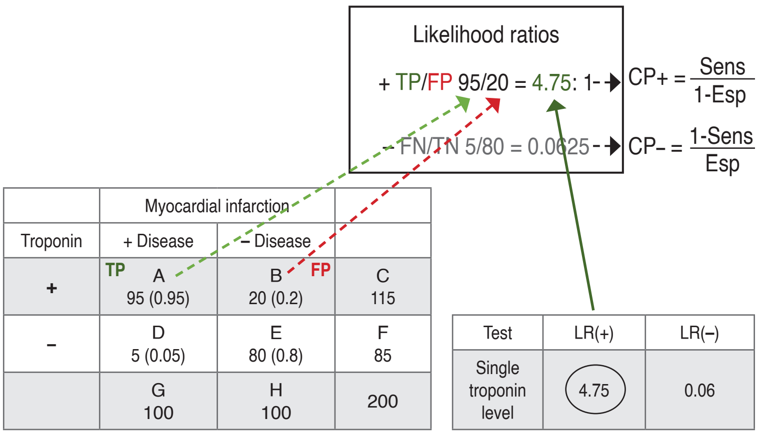

Figure 4: The proportion of true positives and false positives is the probability quotient of a positive test.

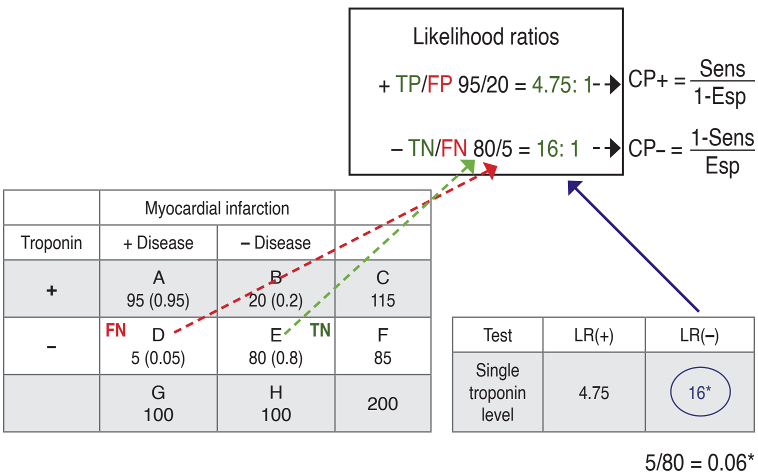

Figure 5: The proportion of true negatives and false negatives is the probability quotient of a negative test.

Let us begin by portraying in a contingency table, the test's results against the «reality» or the best estimation with the gold standard, accordingly to the state of the art *(Table 1)*.

A good test correctly identifies most sick individuals (sensitivity) and the healthy ones (specificity). Now, using the same table, we will calculate the sensitivity, specificity, and predictive values of a positive or negative test *(Figure 2)*. We will utilize the sensitivity and specificity of the troponin test, which is often used to evaluate patients with chest pain *(Figure 3)*. Then, we calculate the Bayes or likelihood ratio for a positive or negative test result. In *Figure 4*, we observe that the ratio of actual positives to false positives is 4.74 to 1. In *Figure 5*, we observe 16 cases of an actual negative test for every false negative, indicating that the test has better sensitivity than specificity and a better predictive value for ruling out the disease than for confirming it.

We will use these likelihood ratios in a nomogram published by Fagan,[4] but in order to illustrate the values for a negative or a positive test on the same graph, since the scale is exponential for the favorable ratio and logarithmic for the negative, we will calculate the negative likelihood ratio as FN/TN instead of TN/FN. In this case, the quotient will be the appropriate fraction (0.0625) for the nomogram *(Figure 6)*.

It has been demonstrated that, although precise numbers are not available because every disease has its own unique conditions, the utility of laboratory and imaging studies is greater in patients with an intermediate probability of having the disease *(Figure 7)*. The next step, as shown in *Table 2* from Diamond & Forrester,[5] is to calculate the prior probability of a disease. The score is built from the risk factors of gender, age, and type of chest pain, as published in the ESC guidelines.[6]

Some critics of Bayesian statistics argue that estimating pretest prevalence can be imprecise. Still, it is preferable to testing in the general population, where the number of false positives and negatives will be high, since no test is 100% accurate. It is easy to calculate that with a sensitivity and specificity of 98%, clearly superior to the tests used in clinical medicine. If you study a population with a 5% prevalence of a disease, most positive and negative results will be false.

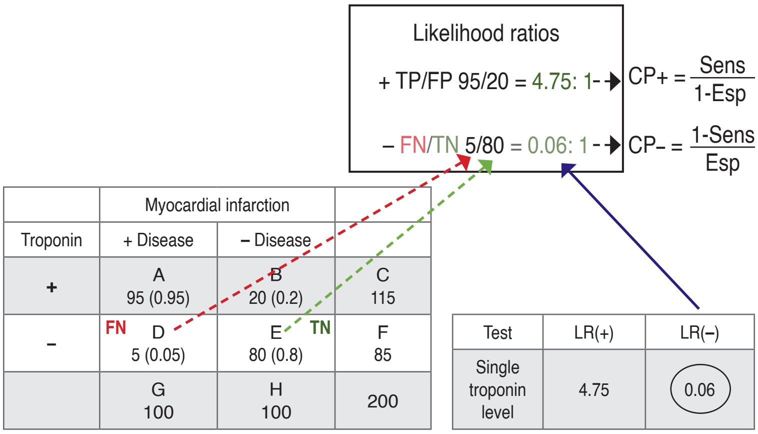

Figure 6: Here, we use the proportion of false negatives to true negatives to plot the predictive value of a positive or negative test in a single nomogram.

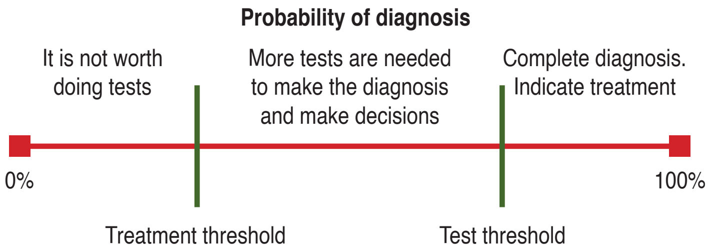

Figure 7: Usefulness of lab and imaging tests according to the prior disease probability.

Then, we will estimate the posterior probability of a disease using the combination of the pretest probability and the likelihood ratios of the test *(Figure 8)*. Using an intermediate probability of 50% for coronary artery disease, we can see that a positive troponin test increases the posterior probability to 80% and a negative one decreases the probability to 5%.[7]

With low and high probabilities, you gain very little from additional tests, as shown in *Figure 9*. Suppose you begin with a low pretest probability (0.1), as many diseases have in the general population. In that case, a positive test will at most put the probability in an intermediate score, not enough to confirm the diagnosis. The same applies to patients with a high pretest probability (0.9), such as those with typical symptoms and risk factors, for whom a negative test result will not necessarily rule out the disease. Therefore, studying patients with an intermediate probability of having a disease yields a larger benefit.

CONCLUSION

The era of reductionism in medicine must come to an end. Physics is the realm of complexity, and biology must be studied in the context of many factors. This results in a larger spectrum of functions than the sum of its parts. Consider the poem "The Blind Men and the Elephant" by John Godfrey Saxe:

Table 2: We can estimate the pretest probability with the Diamond and Forrester pretest probability data.

	High > 70% Typical angor		Intermediate 10-70% Atypical angor		Low < 10% Non angina pain	
Age (years)	M	F	M	F	M	F
30-39	59	28	29	10	18	5
40-49	69	37	38	14	25	8
50-59	77	47	49	20	34	12
60-69	84	58	59	28	44	17
70-79	89	68	69	37	54	24
> 80	93	76	78	47	65	32

It uses three variables: age, sex, and the type of pain, whether it is characteristic of angina pectoris. Numbers show the risk of coronary artery disease.

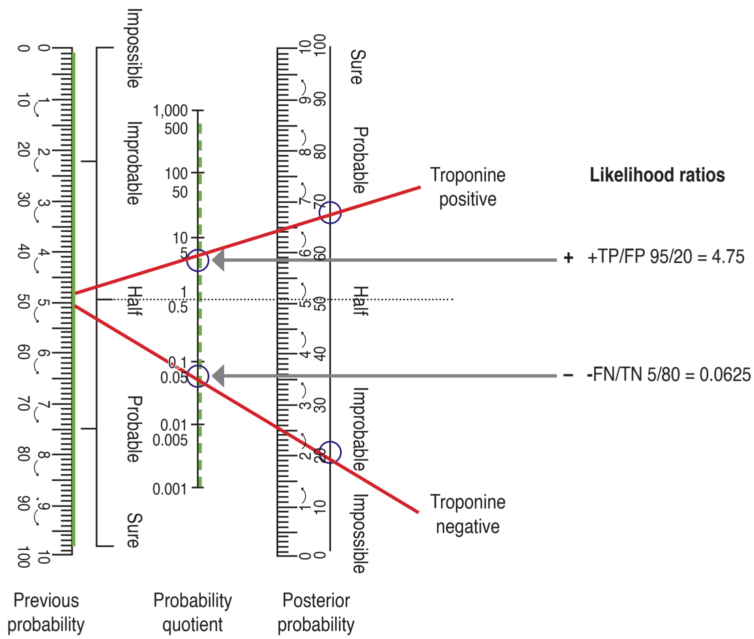

Figure 8: Nomogram to calculate the posterior probability of ischemic heart disease based on the pretest probability and the likelihood ratio of the positive or negative test.

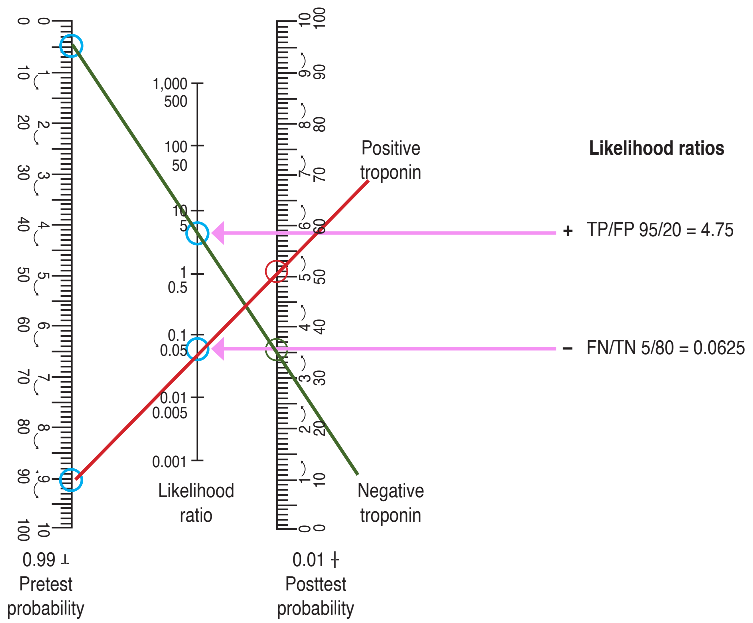

Figure 9: If you start with a low pretest probability (0.1), a positive test result will, at most, place the probability in the intermediate range (green line); similarly, a high pretest probability (0.9) will result in a negative test that does not entirely rule out the disease (red line). Therefore, studying patients with an intermediate probability of having a disease yields a larger benefit.

It was six men of Indostan
To learning much inclined,
Who went to see the Elephant
(Though all of them were blind),
That each by observation
Might satisfy his mind

And so these men of Indostan
Disputed loud and long,
Each in his own opinion
Exceeding stiff and strong,
Though each was partly in the right,
And all were in the wrong!

Each of the men describes what they feel, but no one is able to see the hole!

REFERENCES

1. Uncertainty in medicine. Lancet. 2010; 375 (9727): 1666. doi: 10.1016/S0140-6736(10)60719-2.
2. Moragrega-Adame JL. First study the science, then practice the art: Leonardo da Vinci. Cardiovasc Metab Sci. 2021; 32 (3): 114-116. doi: 10.35366/101303.
3. Stewart I. Does God play dice? The new mathematics of chaos. 2nd ed. London (UK): Penguin Books; 1997.
4. Fagan TJ. Letter: nomogram for Bayes's theorem. N Engl J Med. 1975; 293 (5): 257. doi: 10.1056/NEJM197507312930513.
5. Sorgaard M, Linde JJ, Kofoed KF, Kühl JT, Kelbak H, Nielsen WB et al. Diagnostic value of the updated diamond and Forrester score to predict coronary artery disease in patients with acute-onset chest pain. Cardiology. 2016; 133 (1): 10-17. doi: 10.1159/000438980.
6. Guyatt G, Rennie D, Meade MO, Cook DJ. Users' guides to the medical literature: a manual for evidence-based clinical practice. 3rd ed. New York (NY): McGraw-Hill Education; 2015.
7. Batal O, Malhotra S, Harinstein M, Markowitz J, Hickey G, Agarwal S et al. Performance of traditional pretest probability estimates in stable patients undergoing myocardial perfusion imaging. Circ Cardiovasc Imaging. 2019; 12 (10): e008473. doi: 10.1161/CIRCIMAGING.118.008473.
8. Brush JE Jr, Kaul S, Krumholz HM. Troponin testing for clinicians. J Am Coll Cardiol. 2016; 68 (21): 2365-2375. doi: 10.1016/j.jacc.2016.08.066.

Correspondence:
José Luis Moragrega, MD, FACC
E-mail: jlmoragrega44@yahoo.com.mx