



Revista Colombiana de Cirugía

ISSN: 2011-7582

ISSN: 2619-6107

Asociación Colombiana de Cirugía

Torregroza-Díazgranados, Eduardo de Jesús
Pruebas diagnósticas: Fundamentos de los estudios diagnósticos,
evaluación de la validez e interpretación clínica de sus resultados
Revista Colombiana de Cirugía, vol. 36, núm. 2, 2021, Abril-Junio, pp. 193-204
Asociación Colombiana de Cirugía

DOI: <https://doi.org/10.30944/20117582.716>

Disponible en: <https://www.redalyc.org/articulo.oa?id=355568248003>

- ▶ Cómo citar el artículo
- ▶ Número completo
- ▶ Más información del artículo
- ▶ Página de la revista en redalyc.org

redalyc.org
UAEM

Sistema de Información Científica Redalyc
Red de Revistas Científicas de América Latina y el Caribe, España y Portugal
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto



ARTÍCULO ESPECIAL

Pruebas diagnósticas: Fundamentos de los estudios diagnósticos, evaluación de la validez e interpretación clínica de sus resultados

Diagnostic tests: Fundamentals of diagnostic studies, validity assessment and clinical interpretation of the results

Eduardo de Jesús Torregroza-Díazgranados 

Médico especialista en Cirugía general y subespecialista en Cirugía de Mama y Tejidos blandos; Clínica de Cirugía de Mama, Bogotá, D.C., Colombia.

Resumen

El propósito de esta publicación es describir los fundamentos de los estudios diagnósticos, proporcionando elementos de juicio para evaluar la validez, puntualizar sobre las medidas de utilidad o rendimiento de las mismas y resaltar la importancia de los estudios diagnósticos en la práctica clínica usual. Dentro de los fundamentos de las pruebas diagnósticas se explican los principios básicos sobre las pruebas diagnósticas, la estructura de estas y se detalla cómo se analizan los resultados de la prueba diagnóstica de interés y los resultados de la prueba diagnóstica de referencia. Se evalúa la validez de una prueba diagnóstica examinando tres criterios importantes que sustentan la solidez metodológica de este tipo de estudio: representatividad, constatación y determinación. Por otro lado, se estudia el primer tipo de medidas de rendimiento de una prueba diagnóstica, las medidas de probabilidad (sensibilidad, especificidad, valor predictivo positivo, valor predictivo negativo). Cuando los resultados de una prueba diagnóstica son expresados en una escala cuantitativa continua es necesario escoger “el mejor punto de corte” que se corresponda con una alta sensibilidad y especificidad, para así distinguir los enfermos de los sanos. Con este fin, podemos utilizar las propiedades de la curva COR (acrónimo de Característica Operativa del Receptor). En la presente publicación se describe y pormenoriza la construcción de la curva COR y se puntualiza sobre la mejor manera de analizarla y sacarle provecho. Por último, se ilustra cómo un estudio diagnóstico cambia el paradigma de manejo de una enfermedad.

Palabras claves: diagnóstico; técnicas y procedimientos diagnósticos; prueba de laboratorio; sensibilidad y especificidad; valor predictivo de las pruebas; curva COR.

Fecha de recibido: 13/07/2020 - Fecha de aceptación: 30/09/2020 - Fecha de publicación en línea: 10/02/2021

Correspondencia: Eduardo de Jesús Torregroza-Díazgranados, Calle 111 No. 13-03, Bogotá, D.C., Colombia. Celular: 3175730414

Correo electrónico: torregozad@gmail.com

Citar como: Torregroza-Díazgranados E de J. Pruebas diagnósticas: Fundamentos de los estudios diagnósticos, evaluación de la validez e interpretación clínica de sus resultados. Rev Colomb Cir. 2021;36:193-204. <https://doi.org/10.30944/20117582.716>

Este es un artículo de acceso abierto bajo una Licencia Creative Commons - BY-NC-ND <https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>

Abstract

The purpose of this publication is to describe the fundamentals of diagnostic studies, providing elements of judgment to evaluate their validity, to point out their usefulness or performance measures, and to highlight the importance of diagnostic studies in usual clinical practice. Within the fundamentals of diagnostic tests, the basic principles of diagnostic tests, their structure, and how the results of the diagnostic test of interest and the results of the reference diagnostic test are analyzed and explained. The validity of a diagnostic test is evaluated by examining three important criteria that support the methodological soundness of this type of study: representativeness, verification, and determination. On the other hand, the first type of performance measures of a diagnostic test are studied, the probability measures (sensitivity, specificity, positive predictive value, negative predictive value). When the results of a diagnostic test are expressed on a continuous quantitative scale, it is necessary to choose the best cut-off point that corresponds to high sensitivity and specificity, in order to distinguish the sick from the healthy. To this end, we can use the properties of the ROC (Receiver Operating Characteristics) curve. In this publication, the construction of the ROC curve is described and detailed, and the best way to analyze it and take advantage of it is specified. Finally, it illustrates how a diagnostic study changes the paradigm of disease management.

Keywords: diagnosis; diagnostic techniques and procedures; laboratory test; sensitivity and specificity; predictive value of tests; ROC curve.

Introducción

El diagnóstico puede considerarse como el más importante resultado de la práctica médica, la clave que conduce a un tratamiento específico adecuado. Si el diagnóstico es incorrecto, con mucha probabilidad conllevará a conductas terapéuticas inadecuadas y no exentas de riesgos. Por lo tanto, resulta incomprensible que a pesar del rol primordial que cumplen en la práctica clínica los estudios diagnósticos, el número de publicaciones destinadas a cubrir temas diagnósticos es mucho menor que el de las publicaciones dedicadas a terapias¹.

En términos generales, una prueba diagnóstica se refiere a algún método que permite diferenciar dos o más condiciones, que de otro modo podrían ser confundidas². Por consiguiente, una prueba diagnóstica no sólo es aquella que puede ser verificada en un laboratorio, unidad radiológica, unidad de patología, sino que también incluye aquellos elementos derivados del ejercicio clínico semiológico correcto³.

La mayoría de las veces las pruebas diagnósticas se utilizan para predecir la presencia o

ausencia de una enfermedad; en otros casos, las pruebas diagnósticas se utilizan como métodos para establecer una enfermedad ya conocida; y en otras circunstancias, las pruebas diagnósticas se utilizan para monitorear el curso clínico de una enfermedad^{4,5}.

El propósito de esta publicación es describir los fundamentos de los estudios diagnósticos, proporcionando elementos de juicio para evaluar la validez, puntualizar sobre las medidas de utilidad o rendimiento de las mismas y resaltar la importancia de los estudios diagnósticos en la práctica clínica usual.

Fundamentos de las pruebas diagnósticas

Principios básicos de las pruebas diagnósticas

Todo proceso diagnóstico inicia con la evaluación clínica de los antecedentes personales, los síntomas referidos por el paciente y los hallazgos del examen físico que, tomados en conjunto, arrojan

una aproximación diagnóstica, acerca de una entidad nosológica, la cual llamaremos hipótesis diagnóstica.

Si la impresión diagnóstica no supera el umbral diagnóstico, se rechaza la hipótesis diagnóstica y se descarta el diagnóstico. Si la impresión diagnóstica supera el umbral terapéutico, el médico avala su hipótesis diagnóstica y se inicia el tratamiento correspondiente. Entre el umbral diagnóstico y el umbral terapéutico se ubica la “zona de incertidumbre”, en que el juicio clínico no permite ni descartar ni confirmar el diagnóstico.

Precisamente, es en este escenario, en que las pruebas diagnósticas son de gran provecho, pues mediante su implementación, descartan el diagnóstico, atravesando el umbral diagnóstico o confirman el diagnóstico, pasando el umbral terapéutico⁶ (figura 1).

Estructura de los estudios diagnósticos

Los estudios de pruebas diagnósticas tienen una estructura en común, bien sea que se deriven de experimentos clínicos, estudio de cohorte o estudios transversales, donde los resultados arrojados por la prueba diagnóstica a evaluar se comparan con los resultados de la prueba de referencia, llamada estándar o prueba de oro, que es la mejor prueba diagnóstica disponible en el momento para establecer la presencia o ausencia de una entidad nosológica.

Así, de esta forma, se puede establecer el rendimiento de una prueba diagnóstica, evaluando la concordancia de los resultados de la prueba y los resultados de la prueba de referencia. Entre mayor concordancia exista entre los resultados de estas dos pruebas, mayor es la exactitud de la prueba diagnóstica evaluada⁷. La razón para evaluar una nueva prueba diagnóstica, como sustituta de la prueba alterna o estándar es que, frecuentemente, la prueba de oro es invasiva, costosa o está asociada a complicaciones sustanciales.

La prueba diagnóstica de interés puede expresar sus resultados de tres formas: resultados nominales, ordinales o cuantitativos continuos.

Análisis de los resultados de la prueba diagnóstica de interés y de la prueba diagnóstica de referencia

Los resultados de la prueba diagnóstica de interés y los resultados de la prueba de referencia o estándar de oro se analizan en una *tabla de contingencia*, de donde se derivan las medidas de utilidad o rendimiento de la prueba diagnóstica.

Para el caso de pruebas diagnósticas, una tabla de contingencia se compone de dos vías o entradas y muestra la relación contingente entre las dos pruebas diagnósticas. Las entradas de cada celda son las frecuencias de los resultados de las pruebas. Por convención, la primera entrada, la que ocupa las filas en la tabla de contingencia, corresponde a la prueba diagnóstica que se va a

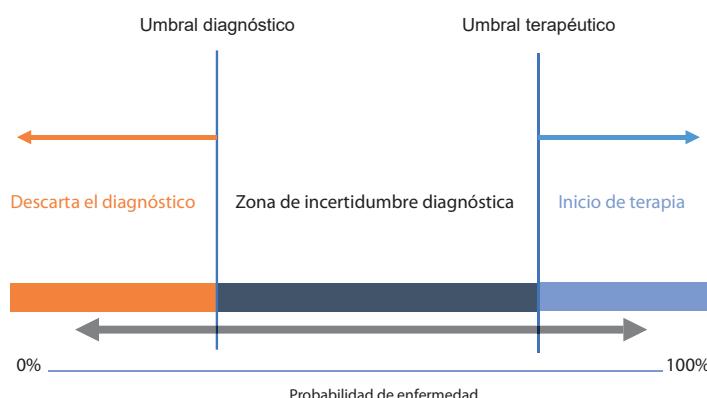


Figura 1. Esquema de umbral diagnóstico y terapéutico

evaluar; y la segunda entrada, la que ocupa las columnas, corresponde a la prueba estándar o de referencia.

La tabla de contingencia, por lo tanto, contiene dos columnas que corresponden al resultado dicotómico de presencia o ausencia de enfermedad, determinada por la prueba estándar. Las filas, por su parte, corresponden a los resultados de la prueba a evaluar.

Las dos columnas y las dos filas forman cuatro celdas, que se designan, por convención, de izquierda a derecha y de arriba abajo con una letra: a, b, c, d (tabla 1). La celda a incluye a los pacientes enfermos en quienes la prueba diagnóstica, correctamente, resultó positiva; *verdaderos positivos*. La celda b incluye a los pacientes sanos en quienes la prueba, equivocadamente, catalogó como positivos; *falsos positivos*. La celda c incluye a los pacientes con la enfermedad en que la prueba diagnóstica, equivocadamente, catalogó como negativos; *falsos negativos*. La celda d incluye a los pacientes sanos en quienes la prueba diagnóstica, correctamente, catalogó como negativos, *verdaderos negativos*.

Tabla 1. Estructura básica de una tabla de contingencia.

Resultado según patrón de referencia			
Resultado de la prueba diagnóstica	Enfermos	No enfermos	Total
Positiva	a	b	a + b
Negativa	c	d	c + d
Total	a + c	b + d	N

Tabla 2. Tabla de contingencia y definiciones.

Prueba o Test diagnóstico		Enfermedad		Total
		Presente	Ausente	
Prueba o Test diagnóstico	Positiva	Verdaderos positivos (VP)	Falsos positivos (FP)	VP + FP
	Negativa	Falsos negativos (FN)	Verdaderos negativos (VN)	FN + VN
	Total	VP + FN	FP + VN	N

En la tabla 2 se pueden apreciar, definiéndolo de otra forma:

- *Verdadero positivo*: el paciente tiene la enfermedad y la prueba es positiva.
- *Falso positivo*: el paciente no tiene la enfermedad y la prueba es positiva.
- *Falso negativo*: el paciente tiene la enfermedad y la prueba es negativa.
- *Verdadero negativo*: el paciente no tiene la enfermedad y la prueba es negativa.

Validez de los estudios diagnósticos

Los resultados de un estudio están determinados, directamente, por la fortaleza de los métodos utilizados para llevar a cabo el estudio⁸. Un estudio diagnóstico bien diseñado debe evitar tres tipos de errores sistemáticos o sesgos: sesgo de representatividad, sesgo de constatación y sesgo de determinación⁹.

Sesgo de representatividad

Un estudio diagnóstico bien diseñado debe incluir pacientes con enfermedades leves a graves y pacientes con otras enfermedades frecuentes con las que se puede confundir.

Sesgo de constatación

A todos los pacientes en quienes se realice la prueba diagnóstica de interés se les debe realizar la prueba diagnóstica de referencia o estándar de oro, sin excepción. De forma incorrecta, cuando la prueba diagnóstica arroja un resultado negativo, los investigadores pueden omitir la realización de la prueba de referencia.

Sesgo de Determinación

Los investigadores que diseñan un estudio diagnóstico deben asegurar que la escogencia de la prueba de referencia es la mejor. Por otro lado, los resultados de la prueba de interés deben ocultarse a la persona que está aplicando e interpretando la prueba de referencia¹⁰.

Medidas de utilidad de las pruebas diagnósticas

Con todos los valores de la tabla de contingencia llenos, se calculan las medidas de utilidad de la prueba diagnóstica. La utilidad o el rendimiento de una prueba diagnóstica se cuantifican mediante dos parámetros o medidas:

1. Medidas de probabilidad.
2. Medidas de razones de probabilidad.

Medidas de probabilidad

Existen dos medidas de probabilidad, sensibilidad y especificidad, que estiman la probabilidad de que los pacientes enfermos y sanos sean diagnosticados como tal por los resultados de la prueba diagnóstica¹¹.

- a. *Sensibilidad*: Probabilidad de que alguien que padece la enfermedad obtenga un resultado positivo en la prueba
- b. *Especificidad*: Probabilidad de que alguien que no padece la enfermedad obtenga un resultado negativo en la prueba.

La sensibilidad y la especificidad se calculan de la siguiente manera:

$$\text{Sensibilidad (S)} = \frac{\text{Verdaderos positivos}}{\text{Verdaderos positivos} + \text{Falsos negativos}} \times 100$$

$$\text{Especificidad (E)} = \frac{\text{Verdaderos negativos}}{\text{Verdaderos negativos} + \text{Falsos positivos}} \times 100$$

Una prueba diagnóstica muy sensible significa que tiene un porcentaje muy bajo de falsos negativos. Por el contrario, una prueba con baja sensi-

bilidad significa que tiene un porcentaje muy alto de falsos negativos. Así, cuando una prueba tiene alta sensibilidad, un resultado negativo, descarta la enfermedad.

Una prueba diagnóstica muy específica significa que tiene un porcentaje muy bajo de falsos positivos. Por el contrario, una prueba con baja especificidad significa que tiene un porcentaje muy alto de falsos positivos. Así, cuando una prueba diagnóstica tiene alta especificidad, un resultado positivo, confirma la enfermedad.

Igualmente, existen dos medidas de probabilidad, valor predictivo positivo y valor predictivo negativo, que estiman la probabilidad de que la prueba diagnóstica arroje el diagnóstico correcto, bien sea que esta resulte positiva o negativa¹².

- a. *Valor Predictivo positivo, VPP*: Probabilidad de que alguien con una prueba positiva tenga la enfermedad.
- b. *Valor Predictivo negativo, VPN*: probabilidad de que alguien con una prueba negativa no tenga la enfermedad.

El valor predictivo positivo y el valor predictivo negativo se calculan de la siguiente manera:

$$VPP = \frac{\text{Verdaderos positivos}}{\text{Verdaderos positivos} + \text{Falsos positivos}} \times 100$$

$$VPN = \frac{\text{Verdaderos negativos}}{\text{Verdaderos negativos} + \text{Falsos negativos}} \times 100$$

Hay que ser muy cautelosos en el cálculo de las medidas de probabilidad. La sensibilidad y la especificidad se calculan en dirección de las columnas en forma vertical, mientras que los valores predictivos positivo y negativo se calculan en dirección de las filas, o sea, en dirección de los renglones en forma horizontal (tabla 3).

Tabla 3. Formas de calcular sensibilidad, especificidad, valor predictivo positivo y valor predictivo negativo.

	Enfermedad +	Enfermedad -	
Test +	Verdadero positivo	Falso positivo	Valor predictivo positivo
Test -	Falso negativo	Verdadero negativo	Valor predictivo negativo
	Sensibilidad	Especificidad	

Interpretación clínica de las pruebas diagnósticas cuyos resultados son expresados en forma dicotómica

La tabla 4 muestra el cálculo de las cuatro medidas de probabilidad: sensibilidad, especificidad, valor predictivo positivo y valor predictivo negativo, con una prevalencia de la enfermedad del 64 %. La prueba diagnóstica de interés expresa sus resultados en forma dicotómica: positiva / negativa.

Tabla 4. Cálculo de las cuatro medidas de probabilidad con una prevalencia de la enfermedad del 64 %. Total de pacientes estudiados: 1000.

Prueba estándar		
Prueba diagnóstica	Enfermos	Sanos
Positiva	608 a	b 43
Negativa	32 c	d 317
Total	640	360

La prevalencia debe ser entendida como la proporción de enfermos, catalogados por la prueba de referencia, dividido entre todos los pacientes del estudio.

La fórmula para calcular la prevalencia es:

$$\text{Prevalencia} = \frac{a + c}{a + b + c + d}$$

- * Prevalencia del 64 %: (640 dividido entre 1000) por 100
- * Sensibilidad: $a / (a + c) = 608 / (608 + 32) \times 100 = 95 \%$
- * Especificidad: $d / (b + d) = 317 / (43 + 317) \times 100 = 88 \%$
- * Valor predictivo positivo: $a / (a + b) = 608 / (608 + 43) \times 100 = 93 \%$
- * Valor predictivo negativo: $d / (c + d) = 317 / (32 + 317) \times 100 = 90 \%$

La sensibilidad y la especificidad son medidas inherentes a la prueba diagnóstica y son robustas, no dependientes de la prevalencia de la enfermedad. Por el contrario, los valores predictivos positivo y negativo si cambian con la prevalencia de la enfermedad.

Cuando la prevalencia de la enfermedad es muy baja (ejemplo, 5 %) el valor predictivo positivo baja y el valor predictivo negativo aumenta, dado que la mayoría de pacientes en el estudio son sanos. Por el contrario, cuando la prevalencia de la enfermedad es alta (ejemplo 95 %) el valor predictivo negativo baja y el valor predictivo positivo aumenta, dado que la mayoría de los pacientes en el estudio están enfermos.

Con el fin de aclarar la relación entre la prevalencia y su impacto en los valores predictivos, tanto positivo como negativo, consideremos la utilidad de una prueba diagnóstica con una prevalencia del 5 % (tabla 5) y del 95 % (tabla 6).

Tabla 5. Cálculo de las cuatro medidas de probabilidad con una prevalencia de la enfermedad del 5 %. Total de pacientes estudiados: 1000.

Prueba estándar		
Prueba diagnóstica	Enfermos	Sanos
Positiva	48 a	b 114
Negativa	2 c	d 836
Total	50	950

* Prevalencia del 5 %: (50 dividido entre 1000) por 100.

* Sensibilidad: $a / (a + c) \times 100 = 95 \%$

* Especificidad: $d / (b + d) \times 100 = 88 \%$

* Valor predictivo positivo: $a / (a + b) \times 100 = 29,6 \%$

* Valor predictivo negativo: $d / (c + d) \times 100 = 99,7\%$

Observe, que la caída de la prevalencia se acompaña de una caída drástica del valor predictivo positivo y el aumento significativo del valor predictivo negativo.

Tabla 6. Cálculo de las cuatro medidas de probabilidad con una prevalencia de la enfermedad del 95 %. Total de pacientes estudiados: 1000.

Prueba estándar		
Prueba diagnóstica	Enfermos	Sanos
Positiva	903 a	b 6
Negativa	47 c	d 44
Total	950	50

- * Prevalencia del 95 %: (950 dividido entre 1000) por 100
- * Sensibilidad: $a / (a + c) \times 100 = 95\%$
- * Especificidad: $d / (b + d) \times 100 = 88\%$
- * Valor predictivo positivo: $a / (a + b) \times 100 = 99.3\%$
- * Valor predictivo negativo: $d / (c + d) \times 100 = 48.4\%$

Igualmente, se debe observar, cómo el aumento de la prevalencia se acompaña de aumento significativo del valor predictivo positivo y la caída drástica del valor predictivo negativo. Obsérvese cómo, al variar la prevalencia de la enfermedad, la sensibilidad y la especificidad se mantienen estables.

Interpretación clínica de las pruebas diagnósticas que expresan sus resultados en una escala cuantitativa continua

La mayoría de las veces las pruebas diagnósticas expresan sus resultados en forma dicotómica (prueba positiva, prueba negativa). Sin embargo, en otras ocasiones, los resultados de una prueba diagnóstica pueden ser expresados en una escala cuantitativa continua. La tabla 7 muestra los resultados de los niveles de creatina quinasa (CK) para el diagnóstico del infarto del miocardio¹³.

En esta última situación es necesario escoger “el mejor punto de corte” que se corresponda con una alta sensibilidad y especificidad, para así distinguir los enfermos de los sanos. Para lograr este cometido, debemos generar una serie de tablas simples, en que cada una de ellas utiliza un punto de corte diferente.

Tabla 7. Resultados de creatina quinasa (CK) para el diagnóstico de infarto del miocardio.

Infarto del miocardio		
Prueba de CK	Presente	Ausente
≥ 280 U/L	97	1
80 a 279 U/L	118	15
40 a 79 U/L	13	26
1 a 39 U/L	2	88
Total	230	130

Cuando el punto de corte se ubica en un nivel de CK ≥ 280 U/L, la celda a contiene a los 97 pacientes con infarto del miocardio cuyos niveles de CK fueron ≥ 280 U/L (verdaderos positivos); la celda b contiene a un (1) paciente quien no tiene infarto del miocardio y tiene niveles de CK ≥ 280 U/L (falso positivo). La celda c contiene 133 pacientes con infarto ($118 + 13 + 2 = 133$) cuyos niveles de CK son menores a 280 U/L (falsos negativos); y la celda d contiene 129 pacientes sin infarto ($15 + 26 + 88 = 129$) cuyos niveles de CK son menores a 280 U/L (verdaderos negativos) (tabla 7.1). La sensibilidad y la especificidad para este punto de corte es de 42 % y 99 %, respectivamente.

Tabla 7.1. Punto de corte de CK ≥ 280 U/L

Infarto del miocardio		
Niveles de CK	Presente	Ausente
≥ 280 U/L	97	a
1 a 279 U/L	133	c
Total	230	1
		129

De esta forma se pueden generar las demás tablas correspondientes para los otros puntos de corte de la prueba de CK (≥ 80 ; ≥ 40 ; ≥ 1).

Para el punto de corte ≥ 80 U/L, la celda a contiene a 215 pacientes con infarto ($118 + 97 = 215$) cuyos niveles de CK son ≥ 80 U/L (verdaderos positivos); la celda b contiene a 16 pacientes sin infarto ($15 + 1 = 16$) quienes tienen niveles de CK mayores a 80 (falsos positivos). La celda c contiene 15 pacientes con infarto ($13 + 2 = 15$) quienes tienen niveles de CK menores a 80 U/L (falsos negativos); y la celda d contiene a 114 pacientes sin infarto ($26 + 88 = 114$) cuyos niveles de CK son menores a 80 U/L (verdaderos negativos) (tabla 7.2). La sensibilidad y la especificidad para este punto de corte es de 93 % y 88 %, respectivamente.

Tabla 7.2. Punto de corte de CK ≥ 80 U/L

Infarto del miocardio		
Niveles de CK	Presente	Ausente
≥ 80 U/L	215	a
1 a 79 U/L	15	c
Total	230	16
		114

Para el punto de corte ≥ 40 U/L, la celda a contiene a 228 pacientes con infarto ($13 + 118 + 97$) cuyos niveles de CK son ≥ 40 U/L (verdaderos positivos); la celda b contiene a 42 pacientes sin infarto ($26 + 15 + 1$) quienes tienen niveles de CK ≥ 40 U/L (falsos positivos). La celda c contiene a 2 pacientes con infarto quienes tienen niveles de CK menores a 40 U/L (falsos negativos); y la celda d contiene a 88 pacientes sin infarto cuyos niveles de CK son menores a 40 U/L (verdaderos negativos) (tabla 7.3). La sensibilidad y la especificidad para este punto de corte es de 99 % y 68 %, respectivamente.

Tabla 7.3. Punto de corte de CK ≥ 40 U/L

Infarto del miocardio				
Niveles de CK	Presente	Ausente		
≥ 40 U/L	228	a	b	42
1 a 39 U/L	2	c	d	88
Total	230			130

Para el punto de corte ≥ 1 U/L, la celda a contiene a 230 pacientes con infarto ($2 + 13 + 118 + 97$) cuyos niveles de CK son ≥ 1 U/L (verdaderos positivos); la celda b contiene a 130 pacientes sin infarto ($88 + 26 + 15 + 1$) quienes tienen niveles de CK ≥ 1 (falsos positivos). La celda c contiene a 0 pacientes con infarto quienes tienen niveles de CK menores a 1 U/L (falsos negativos); y la celda d contiene a 0 pacientes sin infarto cuyos niveles de CK son menores a 1 U/L (verdaderos negativos) (tabla 7.4). La sensibilidad y la especificidad para este punto de corte es de 100 % y 0 %, respectivamente.

Tabla 7.4. Punto de corte de CK ≥ 1 U/L

Infarto del miocardio				
Niveles de CK	Presente	Ausente		
≥ 1 U/L	230	a	b	130
Menor a 1 U/L	0	c	d	0
Total	230			130

La tabla 8 muestra los porcentajes de sensibilidad, especificidad, falsos positivos y falsos negativos para cada punto de corte de CK. En este cuadro, al comparar la sensibilidad y la especificidad para cada punto de corte, nos damos cuenta

de que cuando la sensibilidad aumenta, la especificidad disminuye. Recíprocamente, cuando la especificidad aumenta, la sensibilidad disminuye.

Tabla 8. Reporte detallado de sensibilidad, especificidad, falsos positivos y falsos negativos para cada punto de corte de creatina quinasa (CK).

Punto de corte CK	Sensibilidad	Especificidad	Falsos positivos	Falsos negativos
≥ 1	100 %	0 %	100 %	0 %
≥ 40	99,13 %	67,69 %	32,31 %	66,82 %
≥ 80	93,48 %	87,69 %	12,31 %	6,52 %
≥ 280	42,17 %	99,23 %	0,77 %	57,83 %

También podemos notar que cuando la especificidad aumenta, los falsos positivos disminuyen. Esta relación inversa entre especificidad y falsos positivos es de esperar, dado que el porcentaje de falsos positivos es el complemento de la especificidad. El porcentaje de falsos positivos se calcula:

$$(100 - \text{Especificidad}).$$

Igualmente, podemos darnos cuenta de que cuando la sensibilidad aumenta, los falsos negativos disminuyen. Esta relación inversa entre sensibilidad y falsos negativos es de esperar, dado que el porcentaje de falsos negativos es el complemento de la sensibilidad. El porcentaje de falsos negativos se calcula:

$$(100 - \text{Sensibilidad}).$$

Ahora, ¿cómo podemos establecer el mejor punto de corte del nivel de CK?

En primer lugar, mediante las medidas de resumen global de utilidad de una prueba diagnóstica: *el Índice de exactitud y el Índice de Youden*¹⁴.

Estos índices se calculan para cada punto de corte de la prueba y se valen de ellas para establecer el mejor punto de corte:

1. El Índice de exactitud, se calcula mediante la siguiente fórmula:

$$\frac{\text{Verdaderos positivos} + \text{Verdaderos negativos}}{\text{Verdaderos positivos} + \text{Verdaderos negativos} + \text{Falsos positivos} + \text{Falsos negativos}} \times 100$$

2. El Índice de Youden, se calcula mediante la siguiente fórmula:

$$(\text{Sensibilidad} + \text{Especificidad}) - 100.$$

La tabla 9, muestra el índice de exactitud y el Índice de Youden para cada punto de corte de CK. Acorde a esta tabla, el mejor punto de corte de CK, determinado por el Índice de exactitud y el índice de Youden, es un valor de CK igual o mayor a 80 U/L, que tiene un Índice de exactitud del 91,39 % y un Índice de Youden del 81,17 %. De esta manera unos niveles de CK mayores a 80 U/L predicen la presencia de infarto del miocardio, mientras que unos niveles de CK menores a 80 U/L, descartan la presencia de infarto del miocardio.

Tabla 9. Reporte detallado de sensibilidad, especificidad, índice de exactitud e índice de Youden para cada punto de corte de creatina quinasa (CK).

Punto de corte CK	Sensibilidad	Especificidad	Índice de exactitud	Índice de Youden
≥ 1	100 %	0 %	63,88 %	0 %
≥ 40	99,13 %	67,69 %	87,88 %	66,82 %
≥ 80	93,48 %	87,69 %	91,39 %	81,17 %
≥ 280	42,17 %	99,23 %	62,78 %	41,4 %

En segundo lugar, podemos utilizar las propiedades de la curva COR, acrónimo de Característica Operativa del Receptor (o ROC en inglés, *Receiver Operating Characteristic*) para escoger el mejor punto de corte de CK. La curva COR, es un gráfico que muestra la sensibilidad (en su eje vertical) y el porcentaje de falsos positivos (en el eje horizontal). Los valores en los ejes van desde 0 % hasta el 100 %¹⁵.

Para el caso del eje vertical, donde se ubica la sensibilidad de la prueba, el porcentaje de 0 % se ubica en la esquina inferior izquierda y el 100 % se ubica en la esquina superior izquierda. Para el caso del eje horizontal, donde se ubica el porcentaje de falsos positivos, el porcentaje de 0 % se ubica en la esquina inferior izquierda y el 100 % en la esquina inferior derecha.

La gráfica se configura al ubicar el valor de sensibilidad y el porcentaje de falsos positivos para cada punto de corte de la prueba. La figura 2 muestra la curva COR para los valores de CK en infarto del miocardio.

El primer punto de corte (ubicado en la esquina superior derecha) corresponde al punto de corte ≥ 1 U/L. El segundo punto de corte, de

derecha a izquierda, corresponde al punto de corte ≥ 40 U/L. El tercer punto de corte, el que está más cerca de la esquina superior izquierda, es el punto de corte ≥ 80 U/L. El cuarto punto de corte, que se aproxima a la esquina inferior izquierda, es ≥ 280 U/L.

La curva COR, muestra varias propiedades importantes para establecer el mejor punto de corte de una prueba diagnóstica cuyos resultados son expresados en una escala cuantitativa continua.

La primera propiedad importante de la curva COR es que su esquina superior izquierda indica una prueba diagnóstica perfecta: una sensibilidad del 100 % (todos los pacientes con infarto del miocardio son detectados correctamente, y un porcentaje de falsos positivos de 0% (ningún paciente sin infarto es rotulado falsamente, como enfermo). Por lo tanto, se colige que el punto sobre una curva COR que está más próxima a la esquina superior izquierda es el mejor punto de corte, dado que produce menos porcentaje de error.

Los puntos de corte que se ubican cerca de la esquina inferior izquierda y cerca de la esquina superior derecha nos son deseables porque tienen baja sensibilidad o tienen altos porcentajes de falsos positivos. Por consiguiente, el punto de corte ≥ 80 U/L, es el mejor punto de corte para la prueba de CK, dado que se ubica en la esquina superior izquierda. Obsérvese que mediante ambos métodos se llega a la misma conclusión.

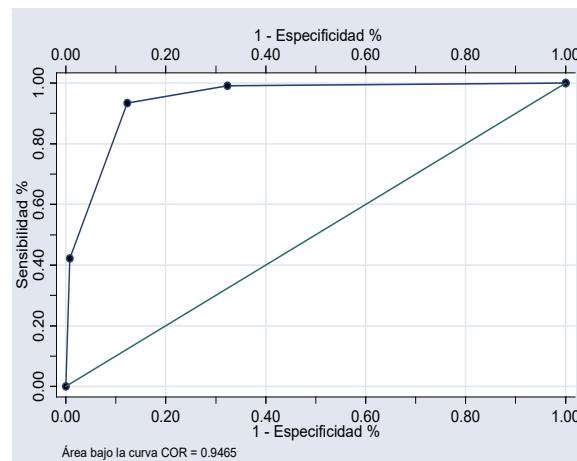


Figura 2. Curva COR, para los valores de creatina quinasa (CK) en infarto del miocardio.

La segunda propiedad importante de la curva COR, es que es un método inestimable para comparar pruebas alternativas para un mismo diagnóstico. La figura 3 compara tres pruebas alternativas para un mismo diagnóstico. Las tres pruebas alternativas son etiquetadas como modalidad diagnóstica 1 (mod1), modalidad diagnóstica 2 (mod2) y modalidad diagnóstica 3 (mod3).

La curva que más se ajuste a la esquina superior izquierda es la que tendrá mayor área bajo la curva COR y por lo tanto tendrá mayor rendimiento diagnóstico. La modalidad diagnóstica 1 tiene un área bajo la curva COR de 0,88, la modalidad diagnóstica 2 tiene un área bajo la curva de 0,93 y la modalidad diagnóstica 3 tiene un área bajo la curva de 0,92. Podemos concluir, rotundamente, que la modalidad diagnóstica 2 (mod2) es la mejor prueba diagnóstica, al tener mayor equilibrio de sensibilidad y especificidad, es decir, mayor área bajo la curva COR.

Importancia de las pruebas diagnósticas en la práctica clínica usual

¿La biopsia de ganglio centinela puede predecir la presencia de metástasis ganglionar en los pacientes con cáncer de seno temprano?

Esta pregunta se resolvió en el año 2003, mediante un experimento clínico llevado a cabo en Italia por Umberto Veronesi, donde se comparó la biopsia de ganglio centinela con el vaciamiento ganglionar axilar en pacientes con cáncer temprano de seno¹⁶.

Antes de la publicación del experimento clínico, el vaciamiento ganglionar era la manera usual de evaluar la presencia de metástasis ganglionares en pacientes con cáncer de seno temprano; sin embargo, el vaciamiento ganglionar se asociaba a una significativa carga de morbilidad y, en un número importante de pacientes, el vaciamiento ganglionar era negativo para metástasis ganglionares, lo que lo constituía en un procedimiento mayor innecesario.

Desde marzo de 1998 hasta diciembre de 1999, 516 pacientes con cáncer de seno temprano (tumores de 2 cm o menos de diámetro y axila clínicamente negativa) fueron aleatorizados en

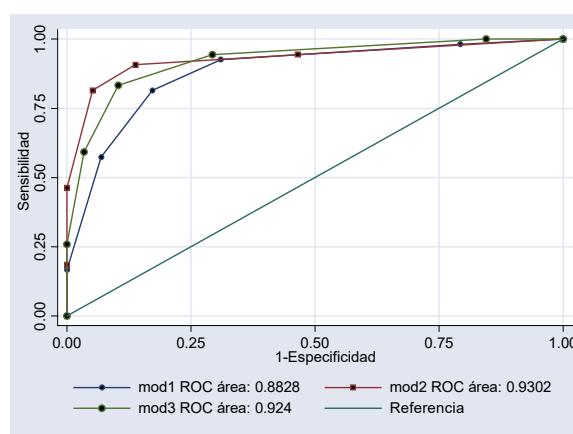


Figura 3. Curva COR, que compara tres pruebas alternativas para un mismo diagnóstico.

dos grupos, uno de biopsia de ganglio centinela y vaciamiento ganglionar, independientemente del resultado de la patología del ganglio centinela (grupo de vaciamiento ganglionar) y otro de biopsia de ganglio centinela, seguida de vaciamiento ganglionar solamente si el ganglio centinela era positivo para metástasis (grupo de ganglio centinela).

Hubo 257 pacientes en el grupo de vaciamiento ganglionar y 259 pacientes en el grupo de ganglio centinela (tabla 10). El ganglio centinela fue positivo en 83 pacientes de 257 en el grupo de vaciamiento ganglionar (32,3 %) y en 92 pacientes de 259 en el grupo de ganglio centinela (35,5 %). Todos los pacientes de ambos grupos, 516 pacientes, fueron seguidos con una mediana de 46 meses.

En el grupo de vaciamiento ganglionar, la disección ganglionar confirmó la presencia de metástasis en 91 pacientes. Entre los 91 pacientes con metástasis ganglionares confirmadas por vaciamiento ganglionar, el ganglio centinela fue positivo en 83

Tabla 10. Resultados del estudio en el grupo de vaciamiento ganglionar: tabla de contingencia.

Vaciamiento ganglionar				
Ganglio centinela	Con metástasis		Sin metástasis	
Positivo	83	a	b	0
Negativo	8	c	d	166
Total	91			166

pacientes; en 8 pacientes el ganglio centinela fue negativo. En 166 pacientes el vaciamiento ganglionar fue negativo para metástasis ganglionares, al igual que la biopsia de ganglio centinela.

Ahora, en primer lugar ¿Son válidos los resultados arrojados por este experimento clínico?

La respuesta es sí. Los investigadores correctamente evaluaron el rendimiento del ganglio centinela como predictor de metástasis ganglionares, en el brazo de vaciamiento ganglionar y no en el brazo de ganglio centinela. ¿Por qué razón? Para evitar el sesgo de constatación. Recuérdese que, en el grupo de vaciamiento ganglionar, a todos los pacientes se les realizó vaciamiento ganglionar independientemente del resultado de la biopsia del ganglio centinela. Por el contrario, en el brazo de ganglio centinela se realizó biopsia de ganglio centinela seguida de vaciamiento ganglionar solamente si el ganglio centinela era positivo para metástasis. En este brazo, si el ganglio centinela era negativo, el paciente no era llevado a vaciamiento ganglionar.

Por otro lado, el experimento clínico efectivamente incluyó un espectro amplio de pacientes con diferentes pronósticos. El 25 % de los pacientes tenía tumores menores o iguales a 1 cm; el 47,9 %, tumores entre 1,1 cm y 1,5 cm; el 26,8 %, tumores entre 1,6 cm y 2 cm; el 31,9 %, tumores grado I; el 46,9 %, tumores grado II, y el 21,3 %, tumores grado III.

Por último, en este experimento clínico no hubo sesgo de determinación, debido a que las personas que realizaron la biopsia del ganglio centinela y el vaciamiento ganglionar (cirujanos) no fueron los que realizaron el estudio de anatomía patológica (patólogos).

En segundo lugar, ¿Cuáles fueron los resultados más importantes de este estudio?

Los resultados del rendimiento de la biopsia del ganglio centinela fueron los siguientes:

* Prevalencia de metástasis del 35,4 %. Es decir, que se realizaron 64,6 % de vaciamientos ganglionares innecesarios.

* Sensibilidad: $a / (a + c) \times 100 = 91,2 \%$.

* Especificidad: $d / (b + d) \times 100 = 100 \%$.

* Valor predictivo positivo: $a / (a + b) \times 100 = 100 \%$.

* Valor predictivo negativo: $d / (c + d) \times 100 = 95,4 \%$.

A los 60 meses de seguimiento, no hubo una diferencia estadística significativa en la sobrevida global entre el grupo de vaciamiento axilar y el grupo de ganglio centinela, $p=0,15$. Así, los resultados de este experimento clínico constituyen un hito en la historia de la cirugía de seno, ya que a partir de este estudio, se realiza la evaluación del estado ganglionar por biopsia de ganglio centinela y no mediante el vaciamiento axilar.

Cumplimiento de normas éticas

Consentimiento informado: Este estudio es una revisión de la literatura, y como tal no hay necesidad de un consentimiento informado ni de aprobación del Comité de Ética Institucional.

Conflictos de interés: Ninguno declarado por el autor.

Fuentes de financiación: Los recursos de financiación del proyecto provienen en su totalidad de aportes del autor.

Referencias

1. Capurro D, Rada G. El proceso diagnóstico. Rev Med Chil. 2007;135:534-8.
<https://doi.org/10.4067/S0034-98872007000400018>
2. Salech F, Mery V, Larrondo F, Rada G. Estudios que evalúan en test diagnóstico: interpretando sus resultados. Rev Méd Chil. 2008;136:1203-8.
<https://doi.org/10.4067/S0034-98872008000900018>
3. Sánchez R, Echeverry J. Aspectos sobre diseño y tamaño de muestra en estudios de pruebas diagnósticas. Rev Fac Med. 2001;49:175-80.
4. Deeks JJ. Systematic reviews of evaluations of diagnostic and screening test. BMJ. 2001;323:157-62.
<https://doi.org/10.1136/bmj.323.7305.157>
5. Ruiz A, Ruiz JG. Fundamentos de Investigación Clínica. Exámenes diagnósticos: Aproximación a su uso racional. Pediatría. 1993;28:111-9.
6. Pauker SG, Kassirer JP. The threshold approach to clinical decision making. N Engl J Med. 1980;302:1109-17.
<https://doi.org/10.1056/NEJM198005153022003>
7. Sackett DL, Haynes RB. Evidence base of clinical diagnosis. The architecture of diagnostic research. BMJ. 2002;324:539-41.
<https://doi.org/10.1136/bmj.324.7336.539>

8. Valenzuela L, Cifuentes L. Validez de estudios de tests diagnósticos. *Rev Med Chil.* 2008;136:401-4.
<https://doi.org/10.4067/S0034-98872008000300018>
9. Straus SE, Glasziou P, Richardson WS, Haynes RB. *Medicina Basada en la Evidencia. Cómo practicar y enseñar MBE.* Quinta edición. Barcelona: Elsevier; 2019. p.153-183.
10. Jaeschke R, Guyatt G, Sackett DL, Bass E, Brill-Edwards P, Browman G, et al. Users' guides to the medical literature: III. How to use an article about a diagnostic test A. Are the results of the study valid? *JAMA.* 1994;271:389-91.
<https://doi.org/10.1001/jama.1994.03510290071040>
11. Altman DG, Bland JM. Diagnostic tests 1: Sensitivity and specificity. *BMJ.* 1994;308:1552.
<https://doi.org/10.1136/bmj.308.6943.1552>
12. Altman DG, Bland JM. Diagnostic tests 2: Predictive values. *BMJ.* 1994;309:102.
<https://doi.org/10.1136/bmj.309.6947.102>
13. Smith AF. Diagnostic value of serum-creatinine-kinase in a coronary care unit. *Lancet.* 1967;2:178-82.
[https://doi.org/10.1016/S0140-6736\(67\)90005-0](https://doi.org/10.1016/S0140-6736(67)90005-0)
14. Youden WJ. Index for rating diagnostic tests. *Cancer.* 1950;3:32-5.
[https://doi.org/10.1002/1097-0142\(1950\)3:1<32::AID-CNCR2820030106>3.0.CO;2-3](https://doi.org/10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3)
15. Altman DG, Bland JM. Diagnostic tests 3: receiver operating characteristic plots. *BMJ.* 1994;309:188.
<https://doi.org/10.1136/bmj.309.6948.188>
16. Veronesi U, Paganelli G, Viale G, Luini A, Zurruda S, Galimberti V, et al. A randomized comparison of sentinel-node biopsy with routine axillary dissection in breast cancer. *N Engl J Med.* 2003;349:546-53.
<https://doi.org/10.1056/NEJMoa012782>