



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO  
FACULTAD DE PSICOLOGÍA  
DIVISIÓN DE ESTUDIOS PROFESIONALES  
Octubre, 2010.



Material elaborado por  
Dra. Corina Cuevas Renaud  
Psic. Amalia Alejo Martínez

### VALIDEZ Y FIABILIDAD DE LAS MEDIDAS DE EXPOSICIÓN Y MEDICIÓN

La **validez** se define como la capacidad de un instrumento (por ejemplo, un cuestionario o una prueba de laboratorio) para medir lo que intenta medir, ésta sólo puede determinarse si existe un procedimiento de referencia, también conocido como *estándar de oro*, el cual, es considerado como un procedimiento definitivo para establecer si alguien tiene la característica de interés. Por ejemplo, una prueba que sólo puede un resultado positivo o negativo (i. e., binario; véase tabla 1).

Para tal situación, es necesario entender dos conceptos importantes como lo son la **sensibilidad** y la **especificidad de una prueba**.

La **sensibilidad de una prueba** responde a las siguientes preguntas:

- ¿Cuántos **resultados positivos** se obtendrán en los individuos enfermos?
- ¿Cuántos casos del **total de casos** en la población estudiada pueden identificarse por el resultado de la prueba?

Tabla 1. Representación de una tabla de contingencia de 2 x 2.

Resultados prueba	Enfermedad		Totales
	Enfermos (+)	Sanos (-)	
Positivos	Verdaderos positivos (a)	Falsos positivos (c)	Positivos a+c
Negativos	Falsos negativos (b)	Verdaderos negativos (d)	Negativos b+d
Total	Enfermos a+b	Sanos c+d	N

Por lo tanto, podemos definir la **sensibilidad de una prueba** como la proporción de los individuos clasificados como positivos por el *estándar de oro* que se identifican correctamente por la prueba en estudio. Lo anterior podemos representarlo de la siguiente forma:

$$\text{Sensibilidad} = \frac{\text{verdaderos positivos}}{\text{total de casos positivos}} = \frac{a}{a+c} = \frac{VP}{VP+FN} \times 100\%$$

donde:

**a**= verdaderos positivos

**a+c**= total de casos positivos (enfermos)

**VP/FN**= verdaderos positivos/falsos negativos

El valor que puede asumir la sensibilidad varía del 0 al 1 (100%), es decir, cuanto más alto es el valor, hay una mejor capacidad en la detección de enfermos por medio de la prueba.

Una sensibilidad baja produce pérdida de casos que pudieran ser tratados, siendo más seria la situación de que a mayor gravedad de padecimiento dejar a pacientes enfermos como fuente de infección en la comunidad representaría un costo alto.

No existe un nivel mágico de sensibilidad que determine que una prueba sea aceptable; por ejemplo, en cáncer de colon 75% (1 de cada 4 retrasará su diagnóstico).

La **especificidad de una prueba** en estudio se refiere a la proporción de los individuos clasificados como negativos por el *estándar de oro* que se identifican correctamente por la prueba en estudio. Este parámetro responde a las siguientes preguntas:

- ¿Cuántos **resultados negativos** en personas sin la enfermedad?
- ¿Cuántos **individuos sanos** se confirmarán por el resultado de la prueba?

$$\text{Especificidad} = \frac{\text{verdaderos negativos}}{\text{total de casos negativos}} = \frac{b}{b+d} = \frac{VN}{VN+FP} \times 100\%$$

donde:

**a**= verdaderos negativos

**a+c**= total de casos negativos (sanos)

**VN/FP**= verdaderos negativos/falsos positivos

Al igual que la sensibilidad, el valor de la especificidad varía del 0 al 1 (100%), lo que significa que cuanto mayor sea el valor mayor capacidad de detección de sujetos sanos por la prueba.

Además de la sensibilidad y la especificidad de una prueba, se han desarrollado otros parámetros para poder determinar qué tanta validez tiene ésta al ser utilizada como prueba diagnóstico. Entre estos parámetros se encuentra el *valor predictivo*. El **valor predictivo positivo** de la prueba responde a las siguientes preguntas:

- ¿De todos los resultados positivos, cuántos se han encontrado en persona que sufren la enfermedad?
- ¿Cuál es la probabilidad en el caso de un resultado positivo de la prueba de que la persona realmente esté enferma?

Este índice se calcula de la siguiente manera:

$$\text{Valor predictivo positivo} = \frac{\text{verdaderos positivos}}{\text{Verdaderos positivos} + \text{falsos positivos}} \times 100$$

El valor obtenido representa la probabilidad de que alguien con un resultado positivo en la prueba en estudio, tenga la característica de interés.

Por otra parte, el **valor predictivo negativo** de la prueba responde las siguientes preguntas:

- ¿Qué proporción de todos los resultados negativos corresponde a personas realmente sanas (sin la enfermedad)?
- ¿Cuál es la probabilidad de que las personas con resultados negativos no tengan la enfermedad?

El cálculo se realiza de la siguiente forma:

$$\text{Valor predictivo negativo} = \frac{\text{verdaderos negativos}}{\text{Verdaderos negativos} + \text{falsos negativos}} \times 100$$

Este valor representa la probabilidad de que alguien con un resultado negativo en la prueba en estudio no tenga la característica de interés.

También es posible obtener el valor predictivo global de la proporción de resultados válidos entre la totalidad de las pruebas efectuadas:

$$\text{VG} = \frac{\text{VP} + \text{VN}}{\text{VP} + \text{FP} + \text{VN} + \text{FN}}$$

El valor predictivo de un resultado depende en gran parte de lo común que sea la enfermedad bajo estudio, esto es, cuando la prevalencia es baja, la obtención de un resultado negativo permitirá descartar con mayor seguridad la enfermedad, teniendo entonces un valor predictivo

negativo mayor. Por el contrario, un resultado positivo no permitirá confirmar el diagnóstico, teniendo entonces un bajo valor predictivo positivo. Respecto a los valores óptimos o esperados que deben asumir los valores predictivos de una prueba, hasta ahora no se han propuesto niveles deseables en particular. Sin embargo, se asume que si el valor es menor al 50%, es más probable que realmente no se encuentre la enfermedad presente, lo significa que a más alto valor, el pronóstico es más valioso.

Es importante tener en cuenta que basar resultados en la prevalencia de una enfermedad a veces resulta complicado, por lo que es útil el uso de otros índices que nos permitan determinar cuánto más probable es un resultado concreto según la ausencia o presencia de enfermedad, por ello, se ha desarrollado el concepto de razón de verosimilitudes, razón de probabilidades o cociente de probabilidades.

- Razón de verosimilitudes positiva o cociente de probabilidades positivo (likelihood ratio of positive test): se calcula dividiendo la probabilidad de un resultado positivo en los pacientes enfermos entre la probabilidad de un resultado positivo entre los sanos. Es decir, el cociente entre la fracción de verdaderos positivos (sensibilidad) y la fracción de falsos positivos (1-especificidad).
  - $RV+= \text{Sensibilidad} / 1 - \text{Especificidad}$

Valores altos indican una mejor capacidad para diagnosticar presencia de la enfermedad.

- Razón de verosimilitudes negativa o cociente de probabilidades negativo (likelihood ratio of negative test): se calcula dividiendo la probabilidad de un resultado negativo en presencia de enfermedad entre la probabilidad de un resultado negativo en ausencia de la misma. Se calcula por lo tanto, como el cociente entre la fracción de falsos negativos (1-sensibilidad) y la fracción de verdaderos negativos (especificidad).
  - $RV-= 1 - \text{Sensibilidad} / \text{Especificidad}$

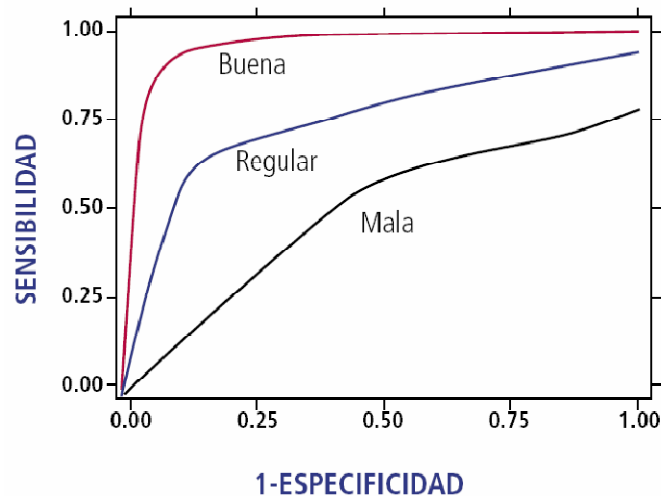
Valores bajos o menores de este parámetro indican una mejor capacidad diagnóstica de la prueba.

Una ventaja importante que poseen estos cocientes respecto a los valores predictivos, es que no dependen de la proporción de enfermos en la muestra, sino de la especificidad y sensibilidad de la prueba.

Por todo lo anterior, podemos decir, que una prueba ideal es aquella que tiene una alta sensibilidad (identifica correctamente a una alta proporción de individuos realmente expuestos o enfermos) y una alta especificidad (da pocos resultados positivos en individuos no expuestos o no enfermos).

El conjunto de pares ordenados 1-Especificidad-sensibilidad (i. e., la sensibilidad en función de los falsos positivos o complemento de la especificidad) obtenidos, se puede representar gráficamente gracias a la curva característica operativa del receptor (ROC, por sus siglas en inglés). En este tipo de curvas se trata el resumen de la validez de una determinación continua para diferentes valores límite (puntos de corte).

El perfil dicotómico respecto a una enfermedad contenido en la representación de los pares ordenados 1-Especificidad-sensibilidad se puede observar en la siguiente gráfica.



**Gráfica 1. Tipos de curva ROC, en función de los pares ordenados 1-especificidad-sensibilidad.**

En la gráfica anterior, se muestran algunas de las pendientes que se pueden obtener así como el valor nominal de cada una de ellas. Cabe mencionar que el criterio para el establecimiento del valor de la gráfica depende de factores como el tipo de enfermedad a estudiar, la experiencia del profesional a cargo, así como de factores inherentes a los sujetos bajo estudio. No obstante, una interpretación generalmente aceptada se lleva a cabo mediante el análisis del área bajo la curva (ABC), que es una medida global de la exactitud diagnóstica. El ABC es definida como la probabilidad de clasificar correctamente un par de individuos (sano y enfermo) seleccionados al

azar de la población, mediante los resultados obtenidos al aplicarles la prueba diagnóstica. Lo anterior podemos representarlo mediante la siguiente expresión:

$$\text{Verdadera ABC ROC} = X_{\text{enfermo}} > X_{\text{sano}}$$

Aunque no existe un punto de corte o *ideal*, por convenio se acepta que el valor del ABC sea mayor o igual a .05. Si no existen diferencias en la distribución de los resultados de la prueba entre los grupos, se toman valores que oscilan alrededor de .05, por ejemplo, valores comprendidos en el intervalo de .05-.07, podrían referir que la exactitud de la prueba es baja. Valores entre .07-.09, pueden resultar útiles sólo bajo ciertas condiciones; en cambio, un valor de .09 en adelante, se da cuando existe una perfecta separación entre las distribuciones, indicando una alta exactitud de la prueba.

La capacidad de discriminación o predictiva de la prueba puede evaluarse estimando el intervalo de confianza del ABC ROC, i. e., el cálculo se obtiene de forma independiente de la prevalencia de la enfermedad en la población de referencia (Fernández y Pértigas, 2003; Molinero, 2002). Esto es, si en el intervalo no se incluye el valor de .05, se dice que la prueba es capaz de discernir entre enfermos y sanos.

Pero, ¿cómo elegimos el valor para establecer un punto de corte para la curva ROC que nos ayude a tomar una decisión respecto a las gráficas obtenidas? A pesar de que la respuesta no es sencilla y tampoco existe una fórmula *absoluta* que nos indique el punto de corte ideal para nuestra investigación, autores como Galen y Gambino (citado en Burgueño, García-Bustos y González-Buitrago, 1995) han sugerido una serie de postulados de tipo práctico respecto a las características de una prueba diagnóstica para elegir el valor del punto de corte a tomar en cuenta.

- 1) Elegir la mayor sensibilidad cuando:
  - a. La enfermedad sea grave y no pueda pasar inadvertida
  - b. La enfermedad sea tratable
  - c. Los resultados falsos positivos no supongan un traumatismo psicológico o económico en los individuos examinados.
- 2) Elegir la mayor especificidad posible cuando:
  - a. La enfermedad sea importante pero difícil de curar o incurable

- b. El hecho de conocer que no se padece la enfermedad tiene una importancia sanitaria y psicológica.
- 3) Debe utilizarse una prueba con un alto valor predictivo positivo cuando el tratamiento de los falsos positivos pueda tener consecuencias graves.
- 4) Un valor global elevado se desea cuando:
- a. La enfermedad sea importante pero curable
  - b. Tanto los falsos positivos como los falsos negativos supongan un traumatismo y éste conlleve consecuencias graves.

A modo de resumen, podemos mencionar tanto la utilidad las curvas ROC como una de las limitaciones que representa el uso de estas curvas. Entre sus utilidades se encuentran las siguientes: a) conocer el rendimiento global de una prueba (ABC), b) comparar dos pruebas o dos puntos de corte. Comparación de dos curvas o de dos puntos sobre una curva, y c) elegir el punto de corte apropiado para *un determinado paciente*.

La principal limitación de este modelo es que sólo contempla dos estados clínicos posibles (sano y enfermo) y no es muy útil su uso para situaciones en que se trata de discernir entre más de dos enfermedades.

Una vez establecido el marco de referencia, es importante aterrizar los conceptos manejados en problemas que nos permitan entender a qué nos referimos. Por ello, a continuación se presentan una serie de ejemplos para que resulte más claro al lector lo ya mencionado.

#### Ejemplo 1

Algunas pruebas proporcionan resultados que no son una simple respuesta si/no. Su lugar, proporcionan valores numéricos en una escala continua de medición. En estas situaciones, se obtiene una alta sensibilidad al precio de una baja especificidad y viceversa. Por ejemplo, cuanto más alta es la presión sanguínea, más probable es la existencia de una enfermedad hipertensiva. Si el nivel diagnóstico o de cribaje para la hipertensión se fija en, por ejemplo, una presión sanguínea diastólica de 90 mmHg, se detectaría a la mayoría de los pacientes hipertensos (alta sensibilidad), pero muchos pacientes no enfermos (con presión sanguínea diastólica superior a 90 mmHg) se clasificarán erróneamente como hipertensos (baja especificidad). Si el nivel de cribaje para la enfermedad hipertensiva se fija a 110 mmHg para la presión sanguínea diastólica, la mayoría de

los individuos no hipertensos se excluirían de la atención posterior (alta especificidad), pero se dejarían de identificar muchos pacientes hipertensos (con presiones sanguíneas diastólicas menores de 110 mmHg, es decir, baja sensibilidad).

En el siguiente ejemplo (2), se ha desarrollado una prueba de laboratorio nueva que mide la concentración de una enzima particular en la sangre y queremos evaluar su valor en el diagnóstico de un cáncer en particular. Esta prueba nueva se aplicó a 360 pacientes hospitalizados y los resultados se compararon con los de un examen anatómo-patológico. Las concentraciones sanguíneas  $\geq 40$  VI se consideraron como valores *positivos*. Los resultados se muestran en la siguiente tabla:

**Tabla 2. Comparación de una prueba de laboratorio nueva con los exámenes anatómo-patológicos en el diagnóstico de un cáncer particular.**

		Examen anatómo-patológico (estándar de oro)		
		Positivo	Negativo	Total
Análisis de sangre	Positivo ( $\geq 40$ VI)	190	80	270
	Negativo ( $< 40$ VI)	0	90	90
	Total	190	170	360

Para la prueba de laboratorio nueva se puede calcular lo siguiente:

- Sensibilidad=  $190/190= 100\%$
- Especificidad=  $90/170= 53\%$

En este ejemplo podrían haberse tomado otros valores de concentración sanguínea como valores límite para definir los resultados del ensayo como *positivos* o *negativos*. En la tabla 3 se presentan los valores de sensibilidad y especificidad del análisis de sangre para diferentes valores límite.

Valor límite (VI)	Resultado del análisis de sangre	Resultado del examen anatómo-patológico	Número de pacientes
40*	+	+	190
	+	-	80
	-	+	0
	-	-	90
80*	+	+	188
	+	-	42
	-	+	2
	-	-	128
120*	+	+	173
	+	-	25
	-	+	17
	-	-	145



280*	+	+	95
	+	-	0
	-	+	95
	-	-	170

Tabla 3. Muestra los resultados obtenidos para diferentes valores límite.

\*Los resultados del análisis de sangre agudos o superiores al valor límite se consideraron positivos.

+= resultado positivo

-= resultado negativo

#### Resultados

- 40 IV: sensibilidad=  $190/190= 100\%$                       especificidad=  $90/170=53\%$
- 80 IV: sensibilidad=  $188/190= 99\%$                       especificidad=  $128/170=75\%$
- 120 IV: sensibilidad=  $173/190= 91\%$                       especificidad=  $145/170= 85\%$
- 280 IV: sensibilidad=  $95/190= 50\%$                       especificidad=  $170/170= 100\%$

En esta tabla se puede apreciar que la sensibilidad de la prueba de laboratorio disminuyó al aumentar el valor límite, mientras que se produjo lo contrario con la especificidad. Lo anterior se ilustra con más claridad en la figura 1.

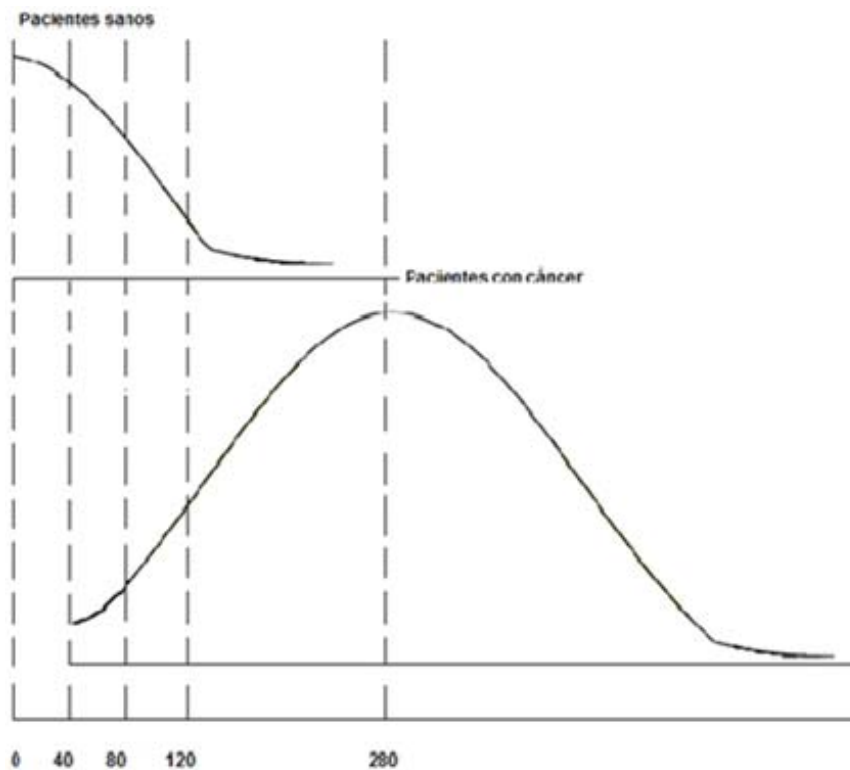
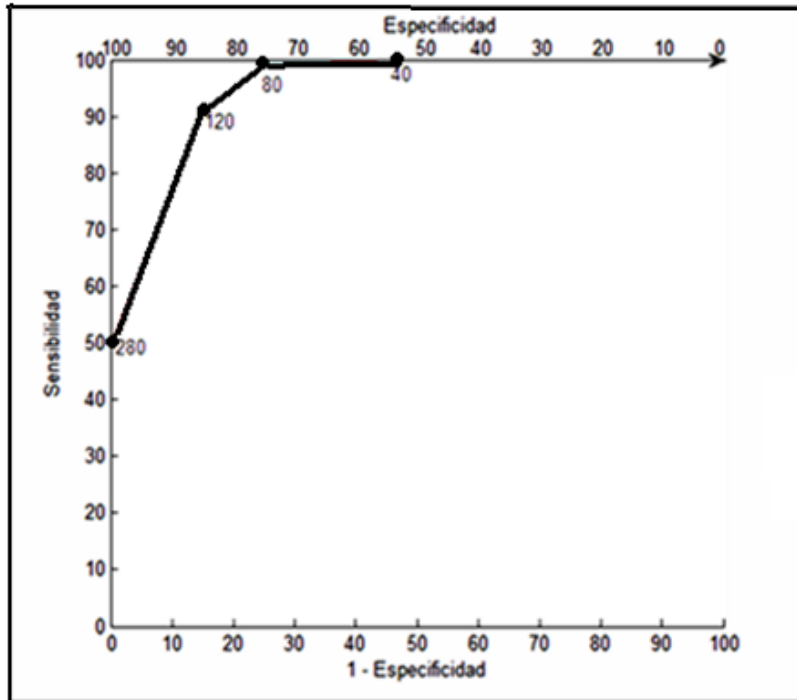


Figura 1. Resultados del análisis de sangre (IV)

Sensibilidad y especificidad del análisis de sangre para valores límite diferentes: datos hipotéticos.

La curva superior describe la distribución del análisis de sangre entre individuos sanos; la curva inferior, la distribución entre pacientes con cáncer (definida por el examen anatómico-patológico). Se utilizaron diferentes valores límite para clasificar los resultados de sangre como *positivos* y *negativos*.

La curva ROC de este ejemplo se muestra a continuación:



Gráfica 2. Curva ROC obtenida al realizar la comparación de oacientes sanos y pacientes con cáncer.

En la gráfica anterior, cuanto más próxima esté la curva ROC de una determinada prueba del extremo superior izquierdo, donde se aumentan la sensibilidad y la especificidad, más válida será la prueba. Una prueba con una curva próxima a la diagonal será inútil y no mejor como técnica diagnóstica que la mera suposición. Para este ejemplo, el punto de corte *ideal* sería de 120, que nos permitirá diferenciar claramente los positivos de los negativos en la población de estudio, para no subestimar o sobreestimar la frecuencia de la enfermedad que se estudia.

### Ejemplo 3

La prueba VIRAPAP tiene sensibilidad y especificidad altas, lo que nos dice que esta prueba es muy válida en la detección de la infección del cuello uterino por el VPH (en comparación con la prueba

de hibridación Southern), y por lo tanto, sus resultados se ven poco afectados por el error de medición (ver tabla 4).

**Tabla 4. Comparación entre los métodos VIRAPAP y de hibridación Southern en el diagnóstico del VPH en el cuello uterino en una muestra de 450 mujeres que asisten a una clínica de enfermedades de transmisión sexual.**

VIRAPAP (prueba nueva)	Southern (prueba estándar de oro)		Total
	Positivos	Negativos	
Positivo	62	22	84
Negativo	7	359	366
<b>Total</b>	<b>69</b>	<b>381</b>	<b>450</b>

Estos datos proporcionan los siguientes resultados para la prueba VIRAPAP:

- Sensibilidad=  $62/69 = .90 \times 100\% = 90\%$
- Especificidad=  $359/381 = .94 \times 100\% = 94\%$
- Valor predictivo de una prueba VIRAPAP positiva=  $62/84 = 74\%$
- Valor predictivo de una prueba VIRAPAP negativa=  $359/366 = 98\%$

Ejemplo 4. La utilidad del ViraPap<sup>®</sup> se comparó con la de la reacción en cadena de la polimerasa (PCR) en 373 mujeres con un diagnóstico reciente de cáncer de cuello uterino. Los resultados se muestran en la siguiente tabla:

		PCR (Prueba estándar de oro)		
		Positivo	Negativo	Total
Prueba ViraPap <sup>®</sup>	<b>Positivo</b>	163	11	174
	<b>Negativo</b>	120	79	199
	<b>Total</b>	<b>283</b>	<b>90</b>	<b>373</b>

**Tabla 5. Comparación de los métodos ViraPap<sup>®</sup> y de la reacción en cadena de la polimerasa (PCR) en la detección de infecciones del cuello uterino por VPH [ Muñoz *et al.* (no publicado)]**

Para la prueba ViraPap<sup>®</sup> se puede calcular lo siguiente:

- Sensibilidad= $163/283=58\%$
- Especificidad= $79/90=88\%$

Como conclusión final, después de presentar varios ejemplos en donde el uso de curvas ROC y la obtención de los parámetros sensibilidad especificidad en casos clínicos, podemos decir que, el análisis económico estará a favor de pruebas con **alta** especificidad, ya que un alto número del resultado de *falsos positivos* determina altos costos en las *pruebas diagnósticas* que puedan llevar a un incremento de los costos del programa. Si el tratamiento es de riesgo para el paciente y se le da sin necesitarlo puede tener como consecuencia efectos secundarios. En suma, las pruebas de baja especificidad pueden producir por iatrogenia efectos indeseables en el paciente y un elevado costo en individuos sin enfermedad.

## Resumen conceptual

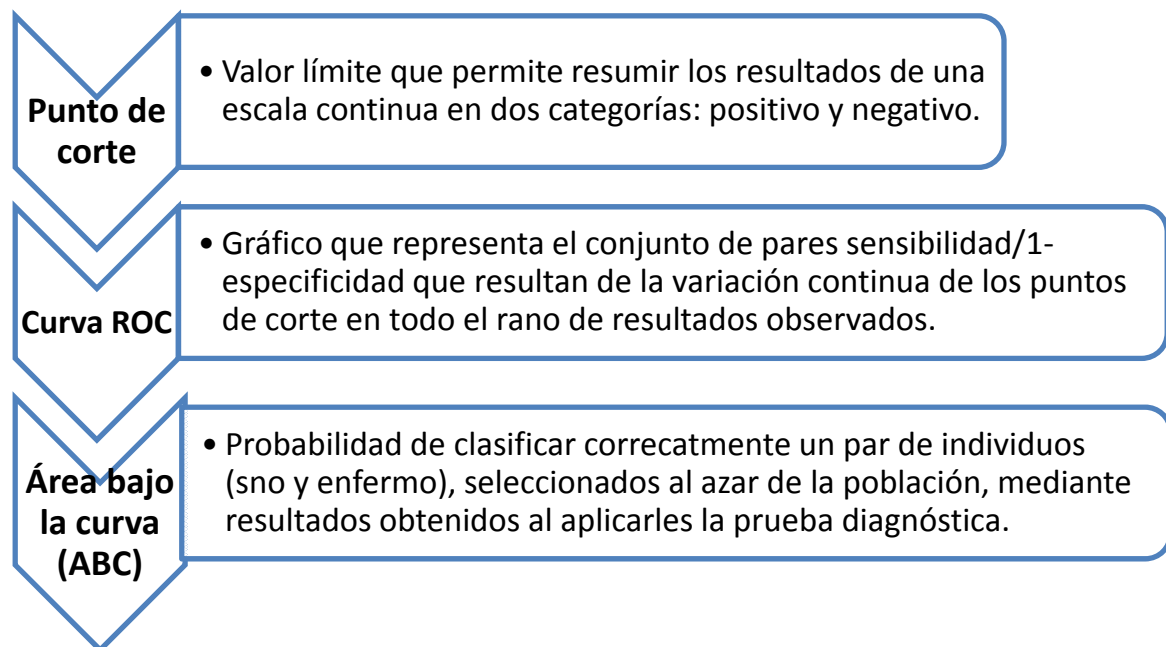
<b>Verdaderos positivos</b>
• La enfermedad está presente y se diagnostica al paciente como enfermo.
<b>Falsos positivos</b>
• La enfermedad <b>no</b> está presente pero al paciente se le diagnostica como enfermo.
<b>Verdaderos negativos</b>
• La enfermedad <b>no</b> está presente y se diagnostica al paciente como sano.
<b>Falsos negativos</b>
• La enfermedad está presente pero no se detectó.

## Sensibilidad

- Proporción de los individuos clasificados como positivos por el *estándar de oro* que se identifican correctamente por la prueba en estudio

## Especificidad

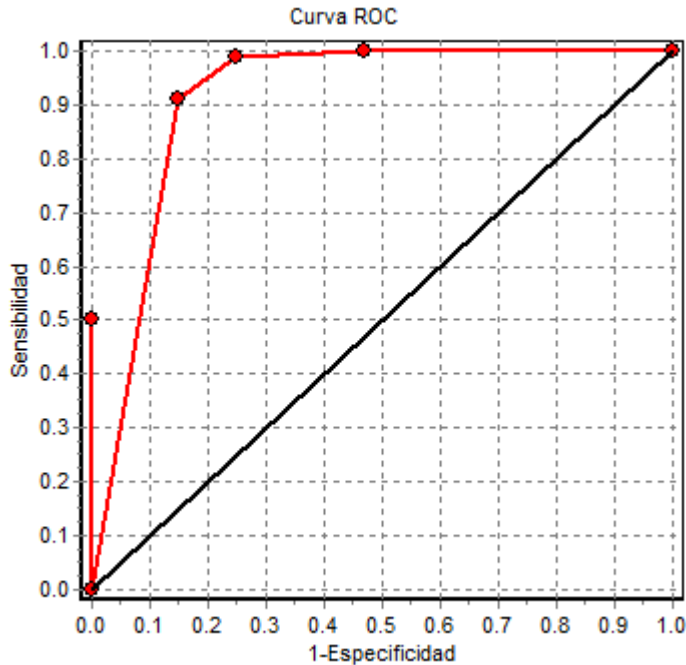
- Proporción de los individuos clasificados como negativos por el *estándar de oro* que se identifican correctamente por la prueba en estudio.



## [2] Curvas ROC simples

Número de tablas: 4  
 Nivel de confianza: 95.0%

Curva ROC obtenida con el programa Epidat 3.1



Área ROC	EE	IC(95%)	
0.9505	0.0092	0.9324	0.9685
	0.0116	0.9278	0.9732
			DeLong
			Hanley & McNeil

## REFERENCIAS

- Burgueño, M. J., García-Bastos, J. L. y González-Buitrago, J. M. (1995). Las curvas ROC en la evaluación de las pruebas diagnósticas. *Med Clin.* 104: 661-670.

Dos Santos SI. *Epidemiología del Cáncer. Principios y Métodos.* Agencia Internacional de Investigación sobre el Cáncer. Lyon Francia 1999 Pp. 11-41

- Fajardo-Gutierrez Et. Al. Utilidad de las curvas de sensibilidad y especificidad conjunta en la aplicación de una prueba diagnóstico. *Salud Pública de México.* 1994; 36: 311-317.

Galen, R. S. y Gambino, S. R. (1975) *Beyond normality: the predictive value and efficiency of medical diagnoses.* New York: J Wiley and Sons, Inc.

López de Ullibarri, I. y Pita, S. F. (1998). Curvas ROC. *Cad Aten Primaria.* 5 (4): 229-235.

Pita, S. F. y Pértegas, S.D. (2003). Pruebas diagnósticas: sensibilidad y especificidad. *Cad Aten Primaria.* 10: 120-124.