

DIPLOMATURA: CONTROL DE CALIDAD EN EL LABORATORIO DE ANÁLISIS CLÍNICOS

CONSEJO DE BIOQUÍMICOS DE LA PROVINCIA DE JUJUY

2025



CONSEJO DE BIOQUIMICOS
DE LA PROVINCIA DE JUJUY

Utilidad de herramientas estadísticas para evaluar una prueba diagnóstica: orientado al laboratorio de inmunología

Bioq. Silvia Barzon

 Consejo de Bioquímicos de Jujuy

DIPLOMATURA CONTROL DE CALIDAD EN EL LABORATORIO DE ANALISIS CLINICOS

MODULO 1: EVALUACIÓN DE MÉTODOS EN EL LABORATORIO DE ANALISIS CLINICOS.
MODULO 2: BASES DE ESTADÍSTICA PARA EL CONTROL DE CALIDAD
MODULO 3: REQUERIMIENTOS DE CALIDAD
MODULO 4: MONITOREO DE DESEMPEÑO ANALÍTICO EN QUÍMICA CLÍNICA
MODULO 5: CONTROL DE CALIDAD INTERNO EN HEMATOLOGÍA Y CITOMETRÍA DE FLUJO
MODULO 6: CONTROL DE CALIDAD INTERNO EN HEMOSTASIA Y BIOLOGÍA MOLECULAR
MODULO 7: CONTROL DE CALIDAD INTERNO EN ENDOCRINOLOGÍA Y MARCADORES TUMORALES
MODULO 8: CONTROL DE CALIDAD INTERNO EN EL LABORATORIO DE MICROBIOLOGÍA
MODULO 9: CONSIDERACIONES ANALÍTICAS DE LAS PRUEBAS RÁPIDAS DE LABORATORIO. LABORATORIO DE URGENCIA

Modalidad Virtual Sincrónica, asincrónicamente por aula virtual
Fechas: 06-08-2025 al 10-12-2025
Acreditación: 130 Hs CÁTEDRA
Horarios de cursado: Miércoles 20.30 a 22 Hs
Informes: comisioncientifica.cbj@gmail.com

 INSCRIPCION <https://forms.gle/xRAy8nwFqctSHFq9>  PROGRAMA ARANCEL

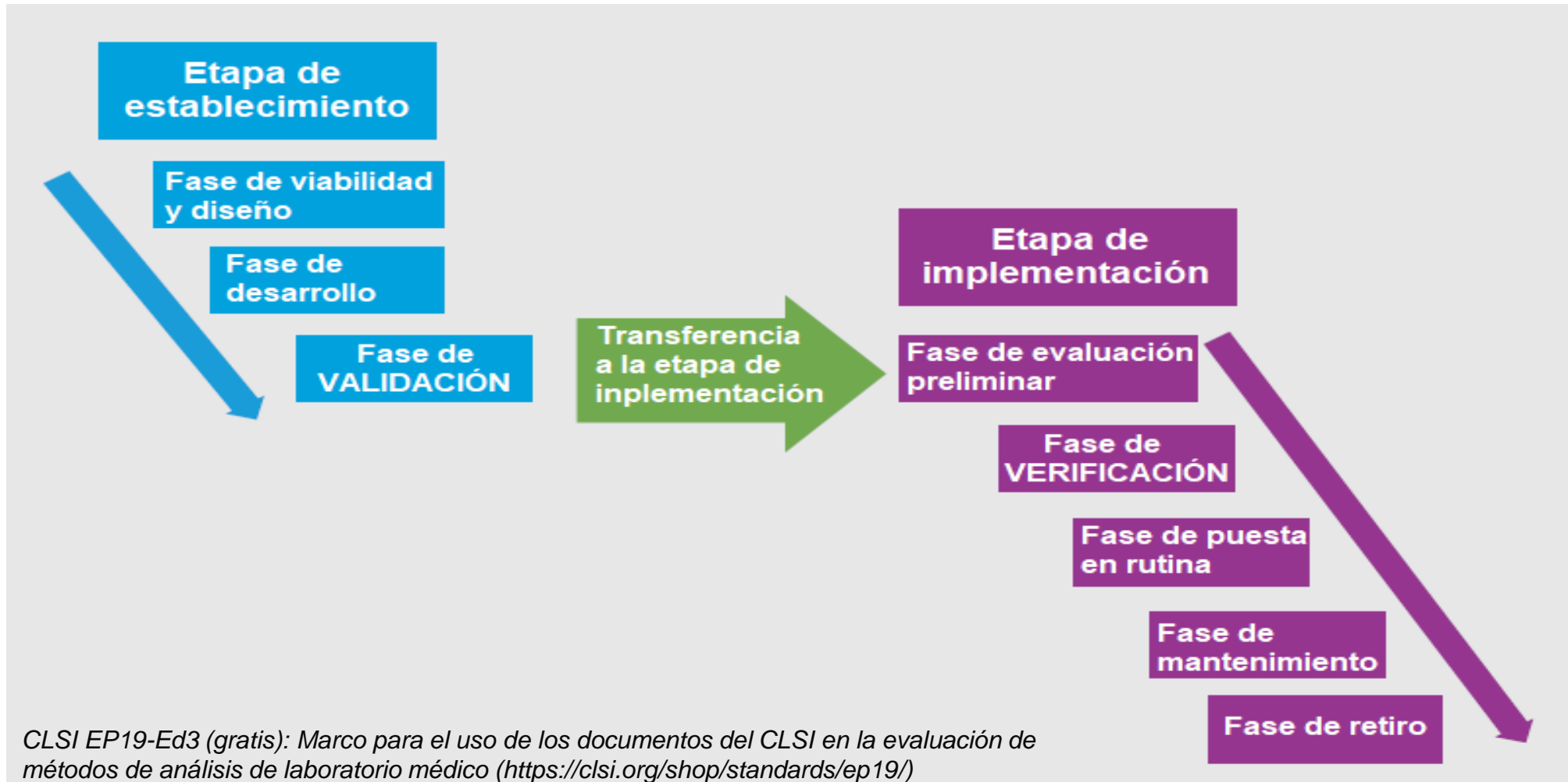
Temario

1. Evaluación de un procedimiento de medida
2. Tipos de resultados obtenidos por una prueba
3. Medidas de desempeño diagnóstico de una prueba (Exactitud Diagnóstica)
 - Pruebas con resultados nominales / categorizados
 - Pruebas con resultados numéricos continuos

Ejemplos orientados al laboratorio de inmunología

Procedimiento de medida: “*modelo de fases del ciclo de vida de una prueba*”

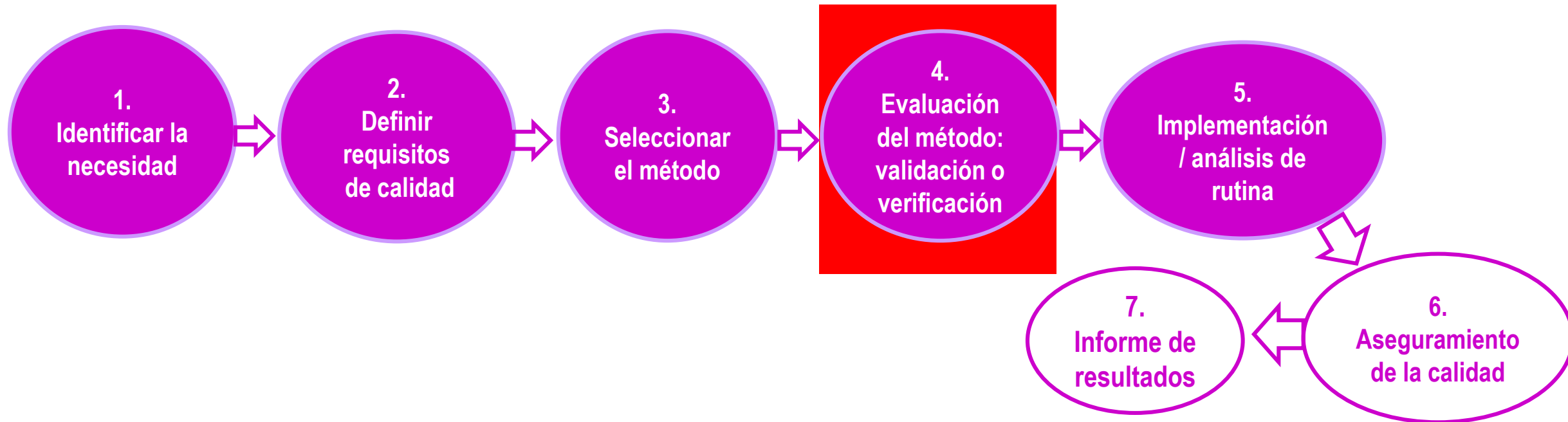
Modelo de fases del ciclo de vida de una prueba: este concepto implica que todos los métodos de prueba pasan por un proceso de **establecimiento** por parte de un “desarrollador” (ej.: fabricante) seguido de una etapa de **implementación** por parte del “usuario final” (ej.: laboratorio), todo de manera secuencial.



Procedimiento de medida: implementación por el laboratorio

Se debe asegurar la calidad analítica de todo procedimiento de medida introducido en el laboratorio, para ello la dirección del laboratorio, junto al personal técnico y de calidad involucrados, deberán evaluar distintos aspectos antes de la implementación.

Pasos a seguir:



Validación y verificación de métodos

VALIDACIÓN

La **validación** establece mediante evidencia objetiva (documentando las características de desempeño) que un método de análisis es adecuado para el uso previsto o intención de uso.

VALIDAR ES ESTABLECER CARACTERÍSTICAS DE DESEMPEÑO ACEPTABLES

Responsabilidad del fabricante o desarrollador

VERIFICACIÓN

La **verificación** es un proceso abreviado que confirma, mediante evidencia objetiva, que un método de análisis **ya validado** es apropiado **para el uso previsto** en el propio laboratorio.

VERIFICAR ES EVALUAR EL DESEMPEÑO DEL MÉTODO EN LAS CONDICIONES DE USO EN EL LABORATORIO

Responsabilidad del usuario

Uso previsto – Intención de uso de una prueba

Uso previsto: uso para el cual está destinado un producto, proceso o servicio de acuerdo con las especificaciones, instrucciones e información proporcionadas por el fabricante. NOTA: el concepto incluye la definición del mensurando, el uso clínico del método de prueba, la detección, el diagnóstico, el pronóstico y/o el monitoreo. (CLSI EP19-Ed3)

Intención de uso / Uso previsto de una prueba



Condiciona:

- Cómo se expresan los resultados (escala de medición)
- El protocolo de validación / verificación que se debe emplear

Tipos de métodos / ensayos

Métodos Cualitativos

Resultados en escala
nominal / binario
(ej.: positivo - negativo)

Protocolo métodos
cualitativos
(ej.: CLSI EP-12)

Resultados en escala
ordinal
(ej.: score / título)

Título
(semicuantitativo)

¿Protocolo?

Métodos Cuantitativos

Resultados en escala
proporcional continua
(ej.: glucemia)

Resultados en escala
proporcional discreta
(ej.: conteo de células)

Protocolo métodos cuantitativos
(ej.: CLSI EP-15A3)

ANA HEp-2 IFI

Métodos semicuantitativos:

- Prueba que tiene un gradiente dosis-respuesta pero para el cual no existe una escala de calibración para determinar la inexactitud e imprecisión. NOTA: Esta definición incluye pruebas con lectura subjetiva de cuantificación (CLSI I/LA02-A2).
- Es esencialmente un método de prueba cualitativo con una opción adicional para un rango de respuesta; podría usar una escala ordinal (CLSI EP19-Ed3)

Verificación de métodos: elección de parámetros a verificar

Parámetros a evaluar	
✓ Veracidad	
✓ Precisión (repetibilidad y precisión intermedia)	
✓ Linealidad (intervalo de medición)	
✓ Límite de detección	
✓ Límite de cuantificación	
✓ Comparación de métodos	
✓ Interferentes	
✓ Estabilidad	
✓ <i>Carry-over</i>	
✓ Robustez	
✓ Incertidumbre de los resultados	
✓ Establecimiento/verificación de intervalos de referencia /valor de corte	
✓ Comparabilidad de resultados	
✓ Sensibilidad y especificidad diagnóstica (exactitud diagnóstica)	✓

Evaluación de un procedimiento de medida

Evaluación analítica (desempeño analítico)

- Precisión / Veracidad
- Especificidad
- Linealidad
- Intervalo de medición
- Límite de detección
- Límite de cuantificación
- Interferentes
- *Carry over*

La evaluación está relacionada al
desempeño del método

Evaluación diagnóstica (Exactitud diagnóstica)

- Sensibilidad diagnóstica
- Especificidad diagnóstica
- Valor predictivo positivo
- Valor predictivo negativo
- *Likelihood Ratio* positivo
- *Likelihood Ratio* negativo
- Curvas ROC

La evaluación está relacionada a una
condición o evento

Pruebas Diagnósticas

- Las pruebas diagnósticas se emplean, en la mayoría de los casos, para predecir la presencia o ausencia de una enfermedad, así como para estadificar una patología ya conocida o monitorear su evolución clínica.
- La solicitud de pruebas diagnósticas abarca estudios realizados por los servicios de imágenes, anatomía patológica y laboratorio, entre otros, e incluye también los elementos que surgen del ejercicio clínico y del examen semiológico.
- En determinadas circunstancias, el examen físico y la anamnesis del paciente pueden ser suficientes para establecer un diagnóstico o iniciar un tratamiento, sin requerir estudios complementarios

Una prueba diagnóstica se define como cualquier método que permita diferenciar, en un individuo, entre dos o más condiciones que de otro modo podrían ser confundidas.

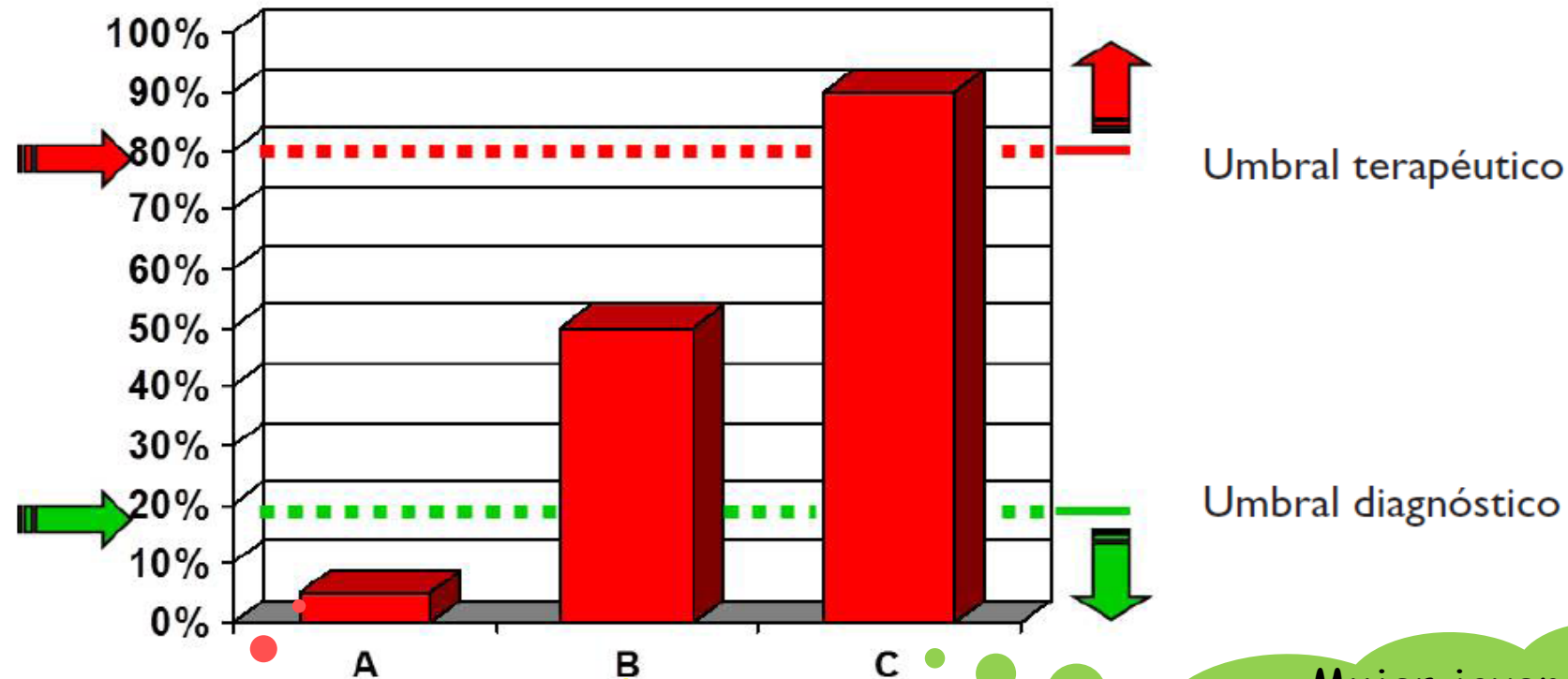
Umbrales Diagnósticos

Todo proceso diagnóstico comienza con la evaluación clínica, que incluye los antecedentes personales, los síntomas referidos por el paciente y los hallazgos del examen físico, a partir de los cuales se establece una impresión diagnóstica inicial.



Puede afirmarse que una prueba diagnóstica resulta "justificable" cuando su resultado tiene el potencial de modificar la conducta clínica o la decisión terapéutica que se habría adoptado en ausencia de dicha prueba.

Umbrales Diagnósticos



Mujer joven con
caída de cabello y
poliartralgiás

Mujer joven con
fotosensibilidad
y leucopenia

Mujer joven con
fotosensibilidad, rash
malar y poliartritis
simétrica; ANA 1/1280
patrón homogéneo (AC-1)

Probabilidades *pre-prueba* y *pos-prueba*

La probabilidad previa a la realización de una prueba diagnóstica (probabilidad pre-prueba) representa la estimación inicial de la probabilidad de que un paciente presente una determinada enfermedad. Esta puede derivarse de diversos elementos clínicos, tales como:

- la presentación clínica del cuadro,
- la presencia de factores de riesgo,
- la experiencia clínica del profesional,
- y la prevalencia de la enfermedad en la población de referencia.

La realización de estudios complementarios permite actualizar dicha estimación, generando una probabilidad posterior a la prueba (probabilidad pos-prueba). Según el resultado obtenido, la probabilidad de confirmar la sospecha diagnóstica puede aumentar, disminuir o, en algunos casos, modificarse mínimamente, lo que indica que la prueba posee escaso valor diagnóstico o resulta no concluyente.

La ventaja de una prueba diagnóstica radica fundamentalmente en poder modificar la probabilidad pre-prueba, permitiendo alejarse de la zona de incertidumbre.

Diseño de estudios de exactitud diagnóstica

- Definir adecuadamente el diseño del estudio, a través de un protocolo o proyecto de trabajo claramente establecido (considerar sesgos).
- Establecer con precisión la metodología empleada para clasificar o determinar el verdadero estado clínico de los individuos.
- Asegurar que la prueba en evaluación no forme parte de los criterios utilizados para la clasificación clínica.
- Considerar todas las variables que puedan influir en los resultados de la prueba, incluidas las fases pre-analítica, analítica y pos-analítica, así como los aspectos metodológicos propios del procedimiento.
- Garantizar que el personal responsable de la realización de la prueba esté adecuadamente entrenado, tanto en el manejo del instrumental, en la técnica empleada como así también en la correcta interpretación de los resultados.

Diseño de estudios de exactitud diagnóstica

Un estudio diagnóstico bien diseñado debe minimizar tres fuentes principales de error sistemático (sesgos):

- **Sesgo de representatividad:** la población incluida debe ser verdaderamente representativa del espectro completo de la enfermedad y de las condiciones relacionadas, evitando selecciones que distorsionen la prevalencia o la severidad del cuadro.
- **Sesgo de verificación (o de constatación):** todos los individuos deben ser evaluados con la prueba de referencia ("*gold standard*"), sin excepciones, para asegurar una clasificación correcta del estado clínico. (Ej.: clasificar correctamente como: "enfermo" / "no enfermo").
- **Sesgo de determinación:** la prueba de referencia debe ser la mejor disponible para establecer el diagnóstico. La realización e interpretación de la prueba en evaluación debe llevarse a cabo de manera ciega respecto de la clasificación clínica de los participantes, a fin de evitar influencias sobre el resultado.

La **exactitud diagnóstica** de una prueba es evaluada por la capacidad de clasificar correctamente a un individuo dentro de dos categorías posibles

Características de una prueba diagnóstica

VALIDEZ: es el grado en que el resultado de la prueba refleja de manera precisa la condición clínica real del paciente. Evalúa qué tan bien una prueba distingue entre quienes tienen y quienes no tienen la enfermedad. Las dos medidas fundamentales de validez son la **sensibilidad y la especificidad diagnóstica**; reflejan la **exactitud intrínseca de la prueba**.

REPRODUCIBILIDAD: la reproducibilidad refleja la capacidad de una prueba para generar resultados consistentes cuando se repite su aplicación en condiciones similares. Puede verse afectada por diversas fuentes de variación:

- Variabilidad biológica, propia del individuo.
- Variabilidad interobservador, atribuible a diferencias entre quienes realizan o interpretan la prueba.
- Variabilidad inherente a la prueba, derivada de la técnica, el equipamiento o los procesos analíticos.

DESEMPEÑO DIAGNÓSTICO: una vez obtenido el resultado, el desempeño de una prueba diagnóstica se refiere a la capacidad de predecir con qué probabilidad un individuo presenta o no una determinada condición o enfermedad. Las principales medidas de desempeño diagnóstico son:

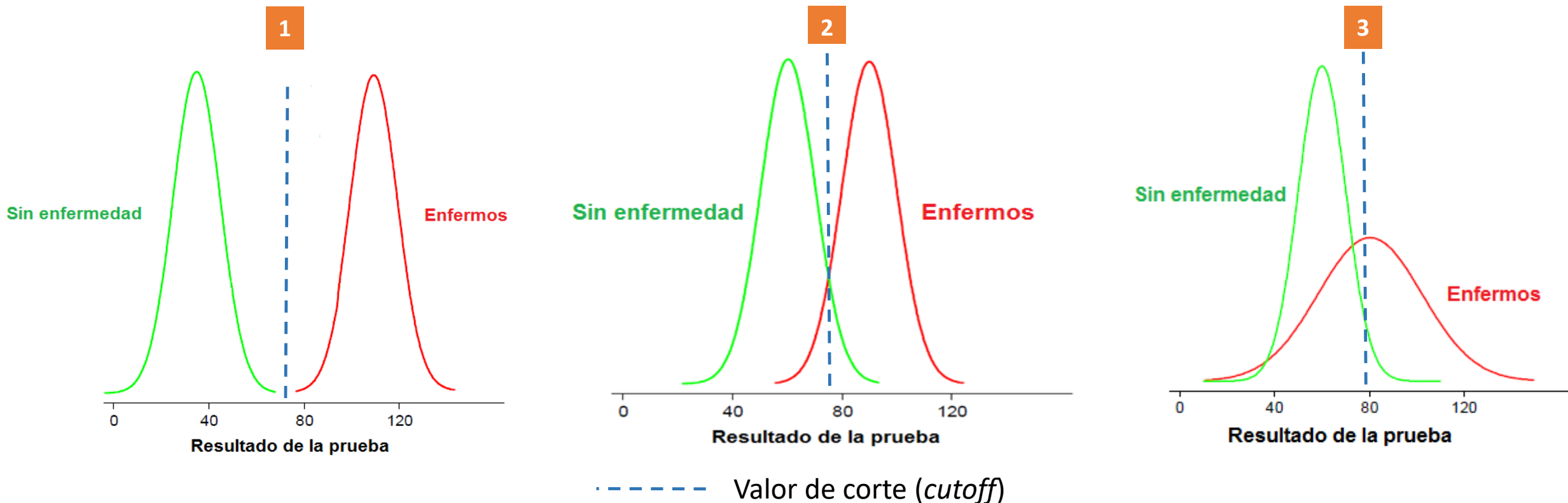
- **Valor predictivo positivo (VPP)**
- **Valor predictivo negativo (VPN)**
- **Razones de verosimilitud (*likelihood ratios*: LR+ y LR-)**

Resultado de la prueba

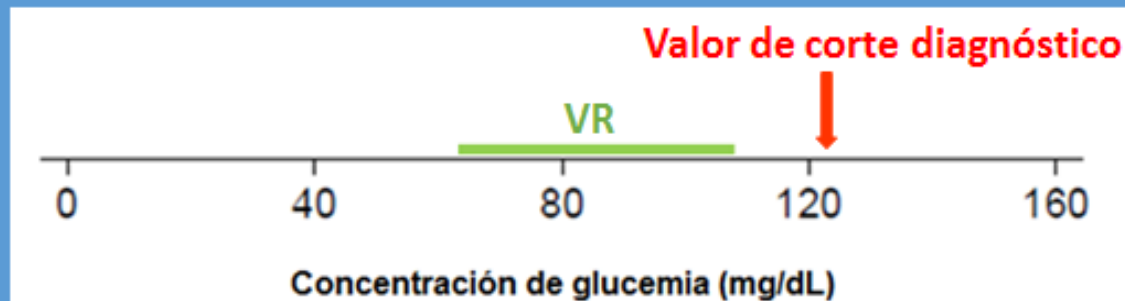
Cuando se realiza una prueba o procedimiento diagnóstico, el resultado puede expresarse de distintas maneras:

- **Mediante un valor de corte (*cut-off*)**, por encima o por debajo del cual se establece la presencia o ausencia de una condición.
- **Mediante la presencia o ausencia de un hallazgo específico** (ej: resultado de una biopsia, IRM, estudio genético)

En determinadas circunstancias, el resultado puede clasificarse como “indeterminado” o “no concluyente”.



Resultado de la prueba



Valor de referencia= 70 – 105 mg/dL

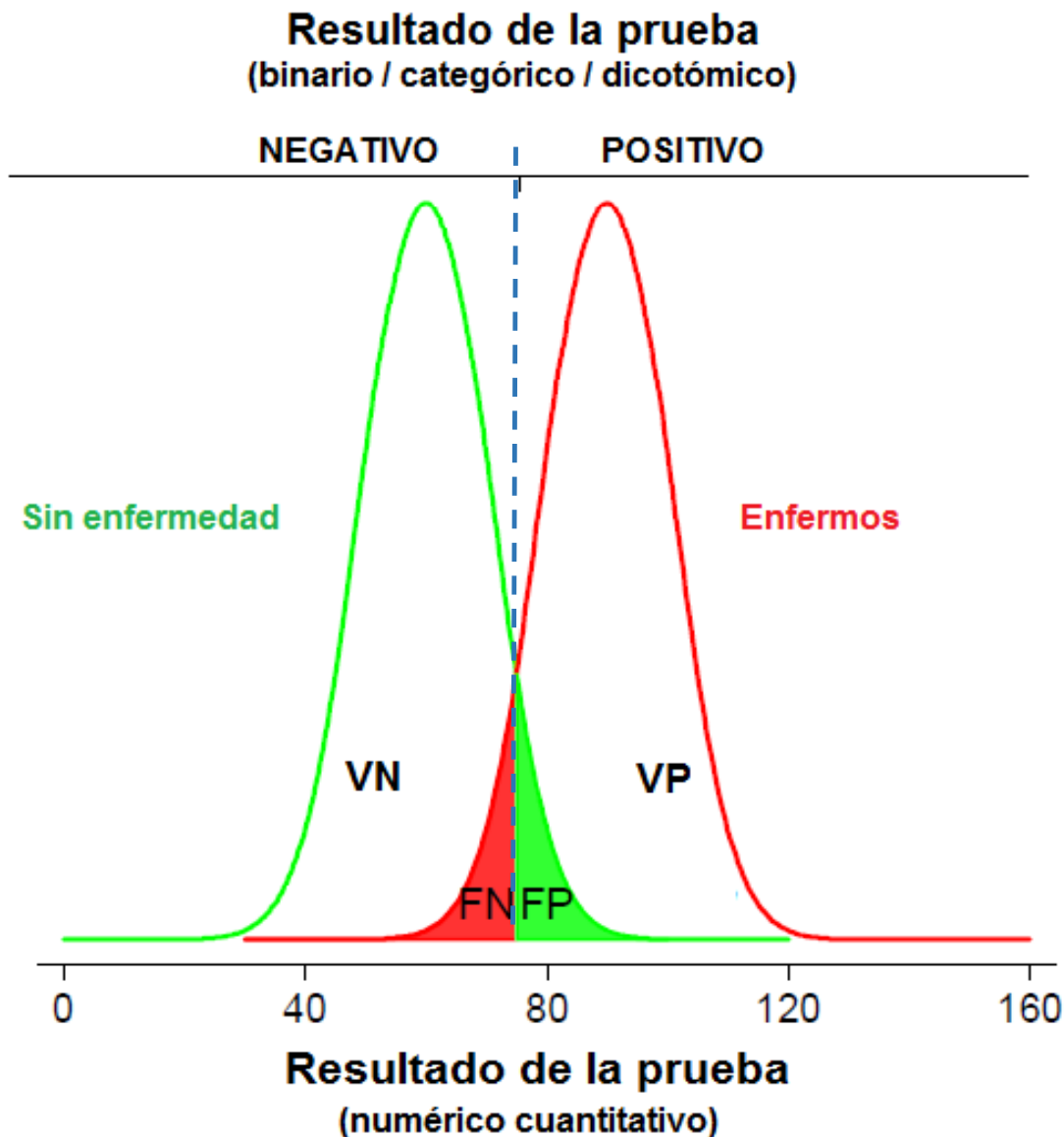
Valor de corte diagnóstico para diabetes= 126 mg/dL

Valor deseable= <100 mg/dL

“Valor deseable” o “Límites de decisión médica” son generalmente recomendados en guías / por consenso.

El valor de referencia es obtenido de “individuos de referencia” de la población; para establecerlos o verificarlos generalmente se emplea la guía del CLSI EP28. El valor de corte diagnóstico se obtiene de estudios de exactitud diagnóstica (ej.: curvas ROC).

Resultado de la prueba



VN: verdadero negativo
FN: falso negativo

VP: verdadero positivo
FP: falso positivo

Medidas de exactitud diagnóstica

A partir de los resultados obtenidos será posible estimar las principales medidas de exactitud diagnóstica, entre ellas:

- Sensibilidad (S) y especificidad (E) diagnóstica.
- Valor predictivo positivo (VPP) y valor predictivo negativo (VPN).
- Razón de verosimilitud positiva (LR+) y razón de verosimilitud negativa (LR–).
- Área bajo la curva (ABC) de la curva ROC.
- Exactitud diagnóstica global.

Es importante considerar que estas medidas dependen, en distinta magnitud, de varios factores:

- La correcta **clasificación de los individuos** según presenten o no la condición estudiada.
- El **espectro de la enfermedad** en la población analizada.
- La **prevalencia de la enfermedad** en dicha población.



El análisis de los datos para obtener las medidas de exactitud diagnóstica depende del tipo de resultado de la prueba a evaluar

Medidas de exactitud diagnóstica: resultado nominal / binario

		Estado de la condición (Ej.: enfermedad)		
Estado de la prueba		Presente	Ausente	
	Prueba +			
	Prueba -			

Tabla 2x2

La **clasificación de los individuos según el estado de la condición debe ser “inequívoca”**. Para ello, es indispensable utilizar el método de referencia o “*gold standard*” disponible, ya sea una prueba de referencia o un criterio consensuado, que permita establecer con la mayor certeza posible la presencia o ausencia de la condición estudiada. En enfermedad “ausente” o “controles”, es recomendable incluir pacientes con enfermedades relacionadas a la condición estudiada (ej.: si se evalúa una nueva prueba en pacientes con Lupus, la muestra de enfermedad “ausente” deberían ser individuos con enfermedades reumáticas autoinmunes sistémicas relacionadas (ej.: AR, Síndrome de Sjögren).

Medidas de exactitud diagnóstica: resultados nominal / binario

		Estado de la enfermedad		
		Presente	Ausente	
Resultado de la prueba	Prueba +	VP	FP	VP + FP
	Prueba -	FN	VN	FN + VN
		VP + FN	FP + VN	VP+FN+FP+VN

Sensibilidad diagnóstica (S): se expresa habitualmente como un porcentaje y representa la proporción de verdaderos positivos (VP) entre el total de individuos que realmente presentan la enfermedad:

$$S = \frac{VP}{VP+FN} \times 100$$

Especificidad diagnóstica (E): se expresa habitualmente como un porcentaje y corresponde a la proporción de verdaderos negativos (VN) entre el total de individuos sin la enfermedad.

$$E = \frac{VN}{FP+VN} \times 100$$

Medidas de exactitud diagnóstica: resultados nominal / binario

		Estado de la enfermedad		
		Presente	Ausente	
Resultado de la prueba	Prueba +	VP	FP	VP + FP
	Prueba -	FN	VN	FN + VN
		VP + FN	FP + VN	VP+FN+FP+VN

Obtengo un resultado positivo ¿con qué seguridad corresponderá a un VP o FP?
Obtengo un resultado negativo ¿con qué seguridad corresponderá a un VN o FN?

Valor predictivo positivo (VPP): se expresa habitualmente como un porcentaje y representa la proporción de verdaderos positivos (VP) entre todos los individuos que obtuvieron un resultado positivo en la prueba.

$$VPP = \frac{VP}{VP+FP} \times 100$$

Valor predictivo negativo (VPN): se expresa habitualmente como un porcentaje y representa la proporción de verdaderos negativos (VN) entre todos los individuos que obtuvieron un resultado negativo en la prueba.

$$VPN = \frac{VN}{VN+FN} \times 100$$

Ejemplo

Equipo comercial para la determinación de Acs anti-CCP por ELISA

Estado de la enfermedad			
	AR (+)	AR (-)	
anti-citrulina (+)	41	9	50
anti-citrulina (-)	17	93	110
	58	102	160

Sensibilidad= $41/58 = 0,71 / 71\%$ (FN= $0,29 / 29\%$)

Especificidad= $93/102 = 0,91 / 91\%$ (FP= $0,09 / 9\%$)

VPP= $41/50 = 0,82 / 82\%$

VPN= $93/110 = 0,85 / 85\%$

Prevalencia= $58/160 = 0,36 / 36\%$

Exactitud diagnóstica global= $\frac{41 + 93}{160} = 0,84 / 84\%$ (porcentaje de pacientes correctamente diagnosticados por la prueba)
(Índice de exactitud)
(Eficiencia diagnóstica)

Interpretación de S, E, VPP y VPN

La sensibilidad diagnóstica: estima la probabilidad de obtener un resultado positivo en quienes realmente presentan la enfermedad; refleja, por lo tanto, la capacidad de la prueba para identificar correctamente a los individuos enfermos.

La especificidad diagnóstica estima la probabilidad de obtener un resultado negativo en quienes no presentan la enfermedad; refleja, por lo tanto, la capacidad de la prueba para identificar correctamente un individuo no enfermo o “saludable”.

El VPP expresa la probabilidad de que un sujeto esté enfermo cuando el resultado de la prueba es positivo.

El VPN expresa la probabilidad de que un sujeto no esté enfermo cuando el resultado de la prueba es negativo.

Dado que tanto el VPP como el VPN dependen de la prevalencia de la enfermedad en la población estudiada, se recomienda cautela al extrapolar estos valores a otras poblaciones o al compararlos con los reportados por otros autores.

Tanto S y E, como los VPP y VPN, pueden expresarse como proporciones (entre 0 y 1) o como porcentajes (entre 0% y 100%). En todos los casos, deben informarse junto a sus respectivos intervalos de confianza del 95% (IC95%).

Teorema de Bayes

Thomas Bayes, clérigo del siglo XVIII, formuló un teorema fundamental para el cálculo de probabilidades condicionales. Aplicado al diagnóstico, permite actualizar la probabilidad pre-prueba de enfermedad cuando se incorpora nueva evidencia proveniente de una prueba.

El teorema establece que el *odds pre-test*, multiplicado por la razón de verosimilitud (LR) de la prueba, produce el *odds pos-test* de enfermedad. De este modo, Bayes describe cómo ajustar la probabilidad inicial a partir del resultado obtenido.

Este enfoque es universalmente válido dentro de la teoría de la probabilidad y explica por qué el cálculo de los VPP y VPN depende de la probabilidad *a priori*, que en la práctica suele aproximarse mediante la prevalencia de la enfermedad en la población estudiada:

$$VPP = \frac{(\text{Sensibilidad}) \times (\text{Prevalencia})}{(\text{Sensibilidad}) \times (\text{Prevalencia}) + (1 - \text{Especificidad}) \times (1 - \text{Prevalencia})}$$

$$VPN = \frac{(\text{Especificidad}) \times (1 - \text{Prevalencia})}{(\text{Especificidad}) \times (1 - \text{Prevalencia}) + (1 - \text{Sensibilidad}) \times (\text{Prevalencia})}$$

Estas fórmulas muestran explícitamente cómo los valores predictivos integran la prevalencia.

VPP, VPN y prevalencia de la enfermedad

Sensibilidad y Especificidad: 90%

Prevalencia %	VPP %	VPN %
99	99,9	8,3
95	99,4	32,1
90	98,8	50
80	97,3	69,2
50	90	90
20	69,2	97,3
10	50	98,8
5	32,1	99,4
1	8,3	99,9

Los valores predictivos dependen fuertemente de la prevalencia de la enfermedad en la población estudiada. A diferencia de la **sensibilidad y la especificidad**, que son **características intrínsecas de la prueba**, el **VPP y el VPN cambian según la frecuencia de la enfermedad**.

- *A mayor prevalencia:* aumenta la probabilidad de que un resultado positivo sea un verdadero positivo (bajos FP)
- *A menor prevalencia:* un resultado negativo tiene más probabilidad de ser un verdadero negativo (bajos FN)

Prevalencia e Incidencia

■ Prevalencia

$$P = \frac{\text{Nº total de casos de la enfermedad}}{\text{población en ese momento}}$$

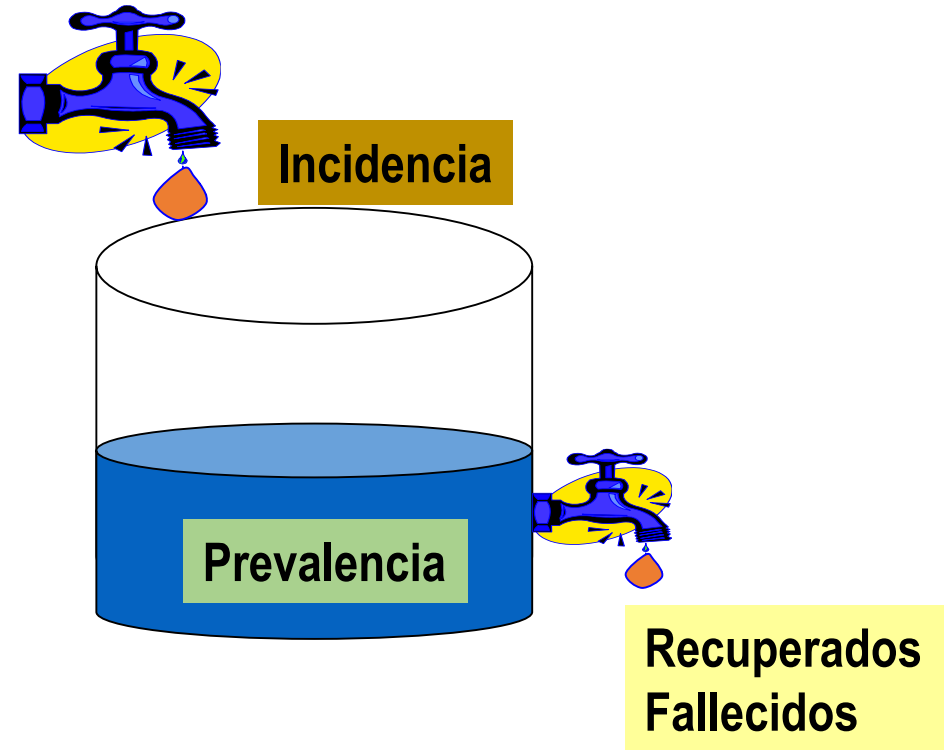
■ Incidencia

Incidencia acumulada: riesgo individual de enfermedad

$$IA = \frac{\text{Nº de casos nuevos en un período de tiempo}}{\text{población en riesgo al inicio de ese período}}$$

Densidad de incidencia: es una tasa; es la velocidad de aparición de un caso nuevo

$$DI = \frac{\text{Nº de casos nuevos}}{\text{Total de tiempo-persona de observación}}$$



A partir de un resultado positivo o negativo ¿cómo evaluamos el desempeño diagnóstico de la prueba para modificar la probabilidad pre-prueba?

¿Qué medida proporciona información sobre la probabilidad de presentar la enfermedad después de realizar la prueba, de manera independiente de la prevalencia?



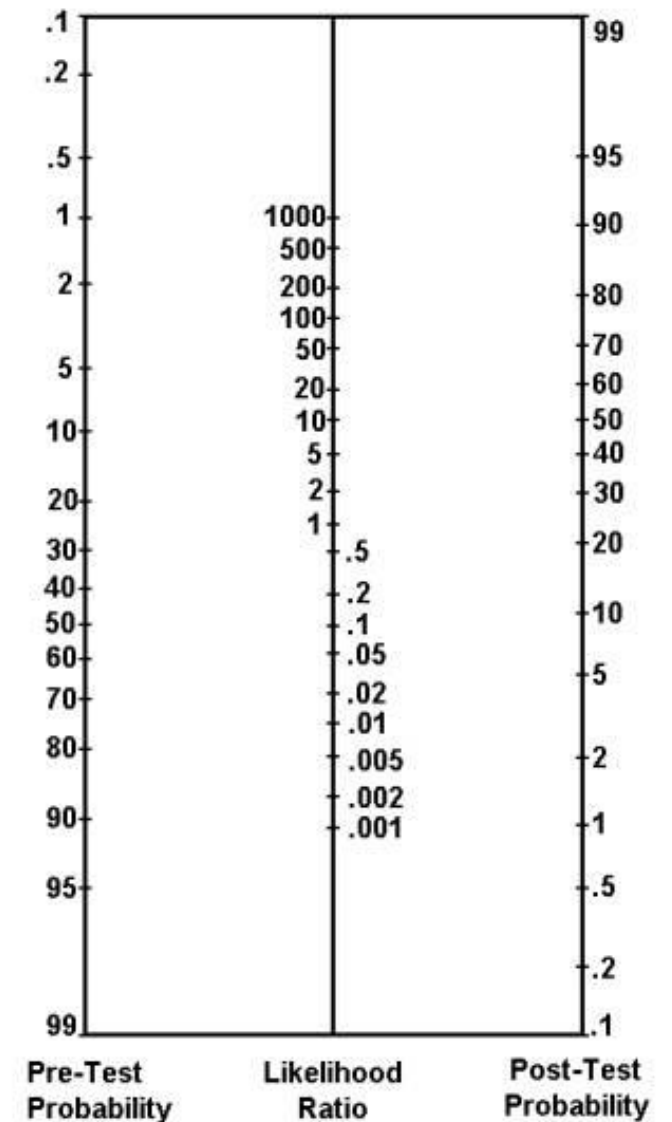
Nomograma de Fagan

Fagan, en 1975, describió un nomograma para graficar el teorema de Bayes, basado en la capacidad de convertir el teorema en una función sumatoria lineal simple.

El nomograma de Fagan consta de 3 columnas:

- La primera es la probabilidad de tener la enfermedad antes de aplicar la prueba, probabilidad *pre-test* o pre-prueba
- La segunda es la razón de verosimilitud o *likelihood ratio* (LR)
- La tercera es la probabilidad *pos-test* o pos-prueba

Algunos nomogramas incluyen S y E para aquellos casos en los que no se cuenta con los LR.



¿Cómo se calculan los LR y cómo se deben interpretar?



Razón de verosimilitud / Razón de probabilidad / *Likelihood ratio (LR)*

Con el fin de dimensionar el beneficio clínico de una prueba diagnóstica, en forma independiente de la prevalencia de la enfermedad, resulta de mayor utilidad la aplicación de los LR.

Los LR son claves en relacionar probabilidades condicionales, en particular la probabilidad *pre-test* para la estimación de la probabilidad *pos-test* (son la expresión del teorema de Bayes).

$$LR = \frac{\text{Probabilidad del resultado de una prueba en pacientes con la enfermedad}}{\text{Probabilidad de ese mismo resultado en pacientes sin la enfermedad}}$$

Los LR indican si el resultado de una prueba diagnóstica, incrementa, disminuye o no modifica la probabilidad *pre-test* de presentar la enfermedad; brindan información sobre la probabilidad de tener o no la enfermedad.

Cada prueba diagnóstica está caracterizada por dos LR:

- LR+ (*likelihood ratio positivo*)
- LR- (*likelihood ratio negativo*)

Likelihood ratio (LR+) y Likelihood ratio (LR-)

LR+: cuan probable es obtener un resultado de la prueba positivo en un individuo con la enfermedad con respecto a uno sin la enfermedad.

$$LR+ = \frac{S}{1 - E} = \frac{\text{Sensibilidad}}{1 - \text{Especificidad}} = \frac{S}{FP}$$

Cuanto mayor sea $LR+ \gg 1$, más fuerte es la evidencia para la presencia de enfermedad, lo que se traduce en un incremento de la probabilidad *pos-test* con respecto a la probabilidad *pre-test*. Los valores varían entre “1” e “infinito”.

LR-: cuan probable es obtener un resultado de la prueba negativo en un individuo con la enfermedad con respecto a uno sin la enfermedad.

$$LR- = \frac{1 - S}{E} = \frac{1 - \text{Sensibilidad}}{\text{Especificidad}} = \frac{FN}{E}$$

Cuanto menor sea $LR- \ll 1$, más fuerte es la evidencia para la ausencia de enfermedad, lo que se traduce en una disminución de la probabilidad *pos-test* con respecto a la probabilidad *pre-test*. Los valores varían entre “1” y “0”

- Los LR+ y LR- no dependen de la prevalencia de la enfermedad
- Los LR son cocientes de proporciones (no son porcentajes)
- Si $LR = 1$ la prueba no es capaz de distinguir entre enfermos y no enfermos (carece de utilidad diagnóstica)

Ejemplo

Estado de la enfermedad

	AR (+)	AR (-)	
anti-citrulina (+)	41	9	50
anti-citrulina (-)	17	93	110
	58	102	160

Sensibilidad= $41/58 = 0,71 / 71\%$ (FN= $0,29 / 29\%$)

Especificidad= $93/102 = 0,91 / 91\%$ (FP= $0,09 / 9\%$)

VPP= $41/50 = 0,82 / 82\%$

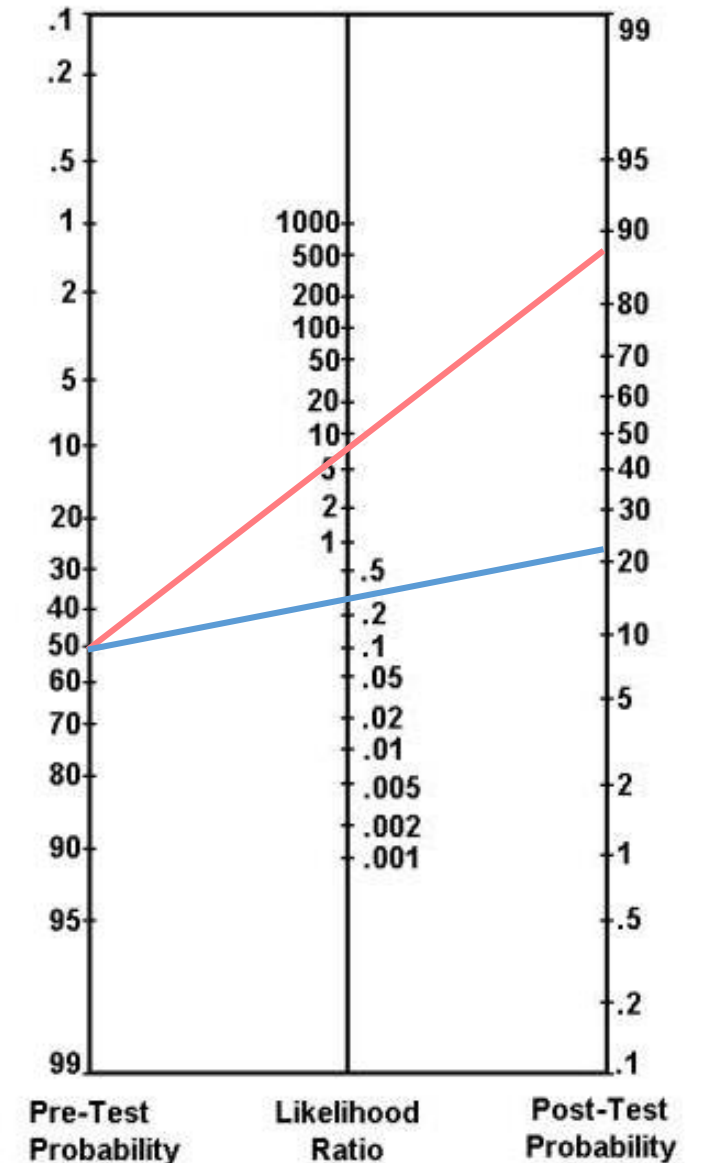
VPN= $93/110 = 0,85 / 85\%$

Prevalencia= $58/160 = 0,36 / 36\%$

Exactitud diagnóstica global= $\frac{41 + 93}{160} = 0,84 / 84\%$

$$LR+ = 0,71 / 0,09 = 7,9$$

$$LR- = 0,29 / 0,91 = 0,3$$



Interpretación de los valores de los *Likelihood ratio*

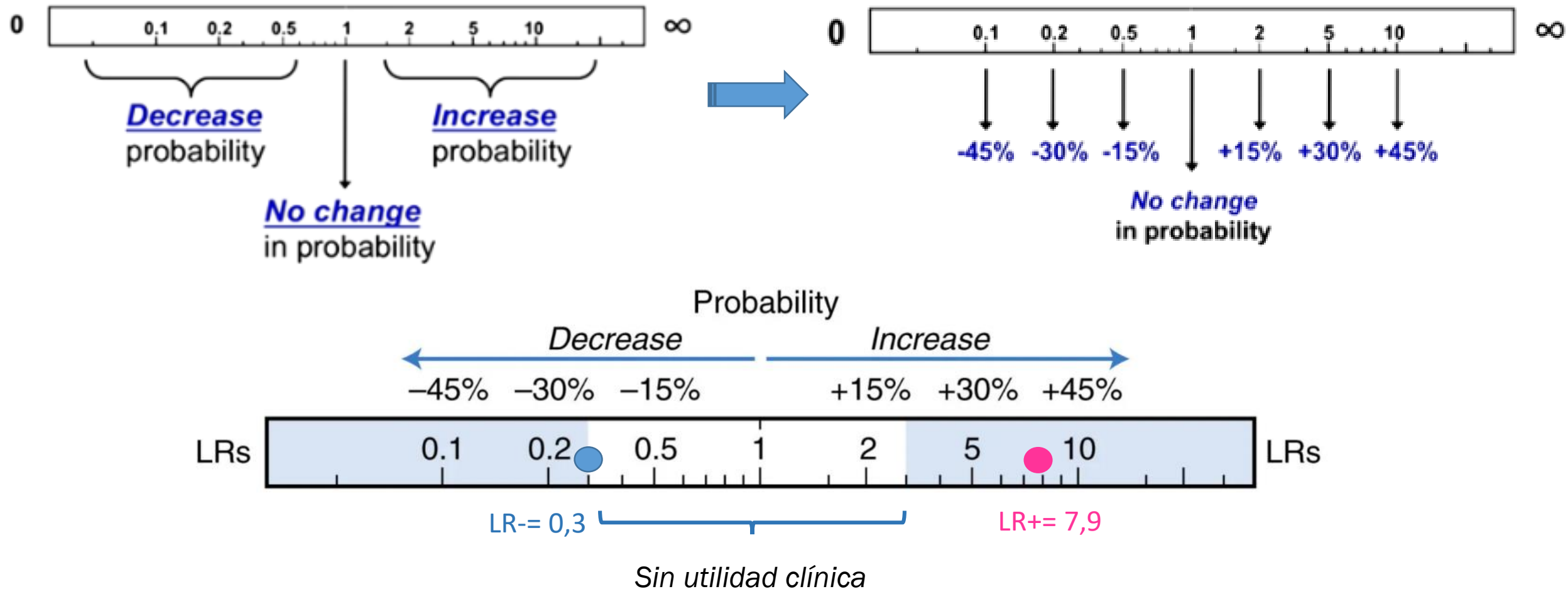
LR+	LR-	Utilidad clínica (cambio de la probabilidad pre-test a pos-test)	Utilidad de la prueba
>10	<0,1	Importante (suficiente para tomar una decisión diagnóstica o terapéutica)	Excelente
7,9 5-10	0,1-0,2	Moderado	Muy útil
<5 y >2	>0,2 y ≤0,5 0,3	Pequeño	Útil
≤2	>0,5	Insignificante	Sin utilidad

Pruebas con LR+ >10 y LR- <0,1 tienen excelente desempeño diagnóstico y generalmente son considerados biomarcadores.

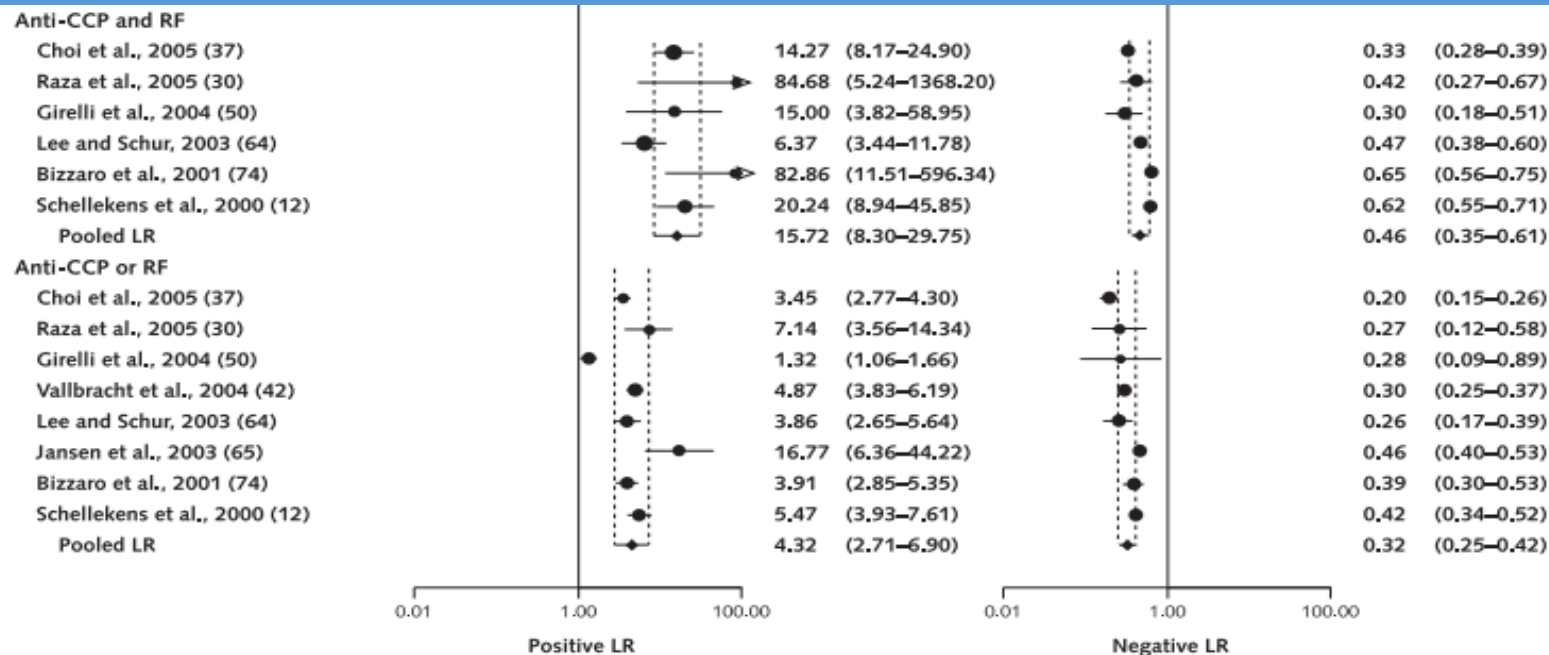
Un LR+ >10 indica que un resultado positivo aumenta fuertemente la probabilidad de enfermedad.

Un LR- <0,1 reduce fuertemente la probabilidad de enfermedad.

Probabilidad pos-prueba según los *Likelihood ratio*



Los LR son herramientas estadísticas que ayudan a estimar cuánto se incrementa ($LR > 1$) o disminuye ($LR < 1$) la probabilidad de padecer la enfermedad o condición evaluada.



Ann Intern Med. 2007;146:797-808.

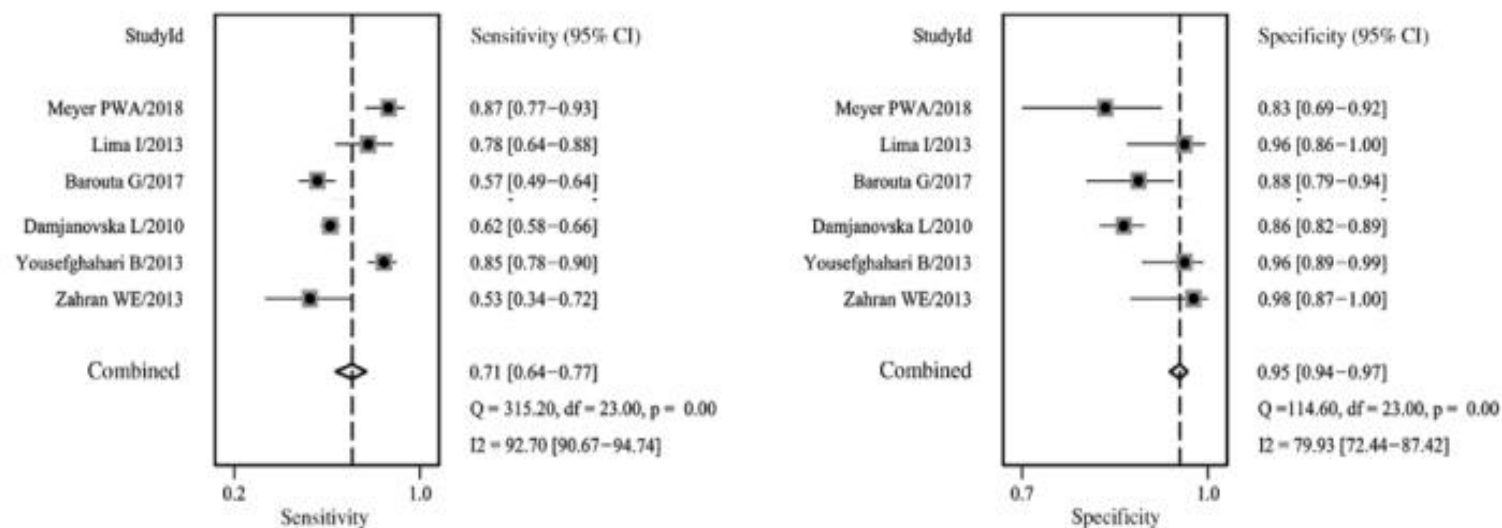


Figure 5: Forest plot of coupled sensitivity and specificity for anti-CCP in the diagnosis of RA vs. DC.

Clin Chem Lab Med 2019; 57(11): 1668–1679

Ejemplo

Autoanticuerpo	Método	Sensibilidad (%)	Especificidad (%)	LR+	LR-
Anti-CCP (AR)	ELISA/CLIA	60–80	90–98	6–40	0,25–0,45
FR (AR)	Nefelometría/ELISA	60–70	70–85	2–4	0,35–0,55
Anti-dsDNA (LES)	Farr/CLIA	50–70	95–100	10–30	0,35–0,50
ANA por IFI	IFI HEp-2	95	45	1,7	0,1
Anti-Sm (LES)	ELISA	25–30	98–99	15–30	0,70–0,75
Anti-Ro/SSA	ELISA/CLIA	60–70	85–95	4–10	0,30–0,45
Anti-Jo1 (miositis)	ELISA	20–30	95–99	5–15	0,70–0,80
AMA-M2	CBP	90–95	95–98	10–20	0,05–0,10
Anti-LKM1	Hepatitis autoinmune 2	70	95	7–14	0,3
Anti-transglutaminasa IgA	ELISA	90–98	95–99	20–60	0,02-0,10
Anti-endomisio	IFI	85–95	>99	>50	0,05-0,15

Variabilidad / incertidumbre de las medidas

Diagnostic test

2x2 table

		Disease		
		Present	Absent	
Test	Positive:	41	9	50
	Negative:	17	93	110
		58	102	

IC95%
↓

Sensitivity	70,69%	57,27% to 81,90%
Specificity	91,18%	83,91% to 95,88%
Positive Likelihood Ratio	8,01	4,20 to 15,28
Negative Likelihood Ratio	0,32	0,21 to 0,48
Disease prevalence	36,25%	28,81% to 44,21%
Positive Predictive Value	82,00%	68,56% to 91,41%
Negative Predictive Value	84,55%	76,41% to 90,73%

Help Test Exit

El **intervalo de confianza** (IC) describe la variabilidad entre el valor obtenido y el verdadero valor en la población de la cual se obtuvo la muestra. Corresponde a un intervalo de valores en el cual se estima que se encuentra, con alta probabilidad (habitualmente IC95%), el verdadero valor poblacional.

Cuanto más estrecho es el IC 95%, mayor es la precisión de la estimación

Ejemplo

Estado de la enfermedad

	AR (+)	AR (-)	
anti-citrulina (+)	41	9	50
anti-citrulina (-)	17	93	110
	58	102	160

Sensibilidad= $41/58 = 0,71 / 71\%$ (FN= $0,29 / 29\%$)

Especificidad= $93/102 = 0,91 / 91\%$ (FP= $0,09 / 9\%$)

VPP= $41/50 = 0,82 / 82\%$

VPN= $93/110 = 0,85 / 85\%$

Prevalencia= $58/160 = 0,36 / 36\%$

Exactitud diagnóstica global= $\frac{41 + 93}{160} = 0,84 / 84\%$

LR+ = $0,71 / 0,09 = 7,9$

LR- = $0,29 / 0,91 = 0,3$

En la muestra de pacientes estudiada, la prueba para determinar Acs anti-CCP por ELISA evidenció:

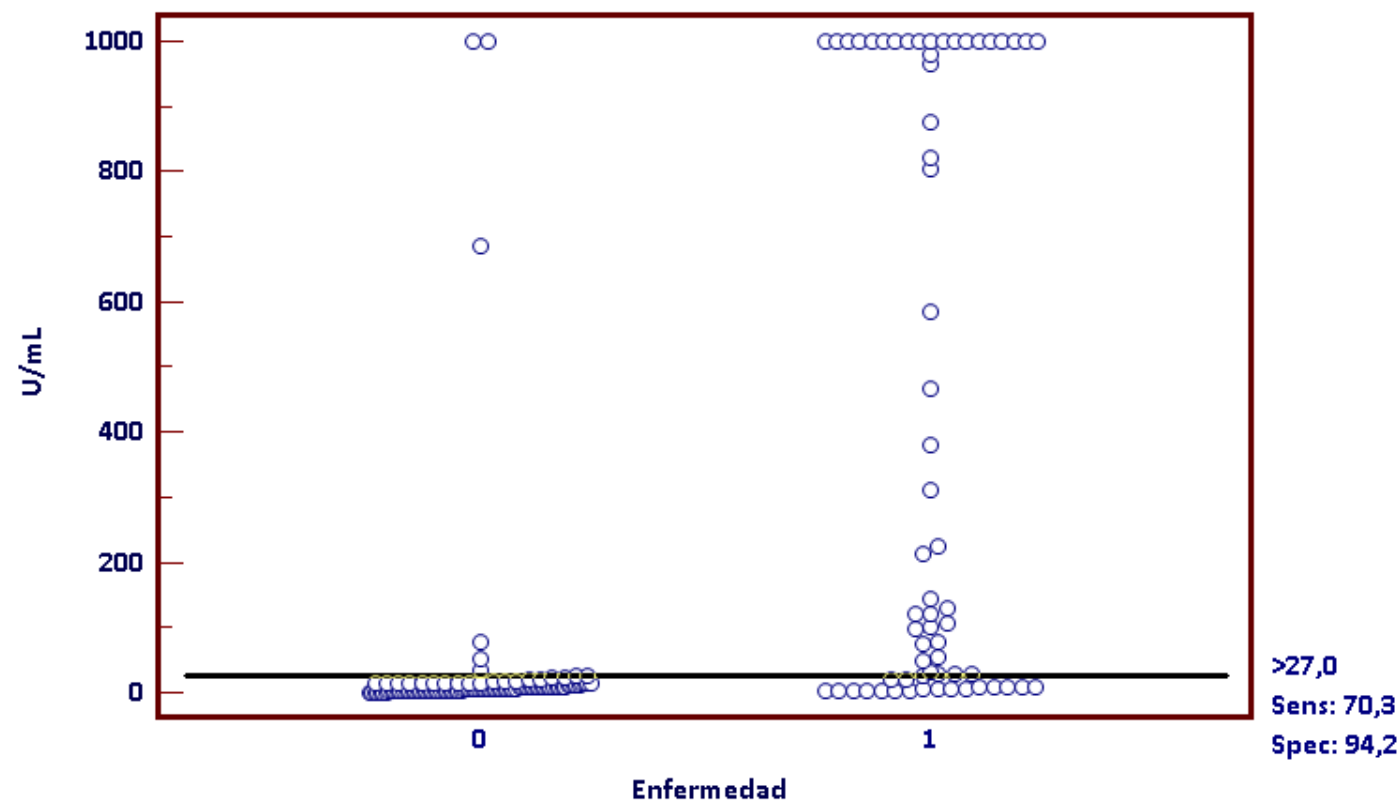
- Alta especificidad (91%)
- Un resultado positivo es clínicamente más útil (LR+= 7,9 “moderada utilidad”)
- Un resultado negativo no descarta AR (LR-= 0,3)



El análisis de los datos para obtener las medidas de exactitud diagnóstica depende del tipo de resultado de la prueba a evaluar

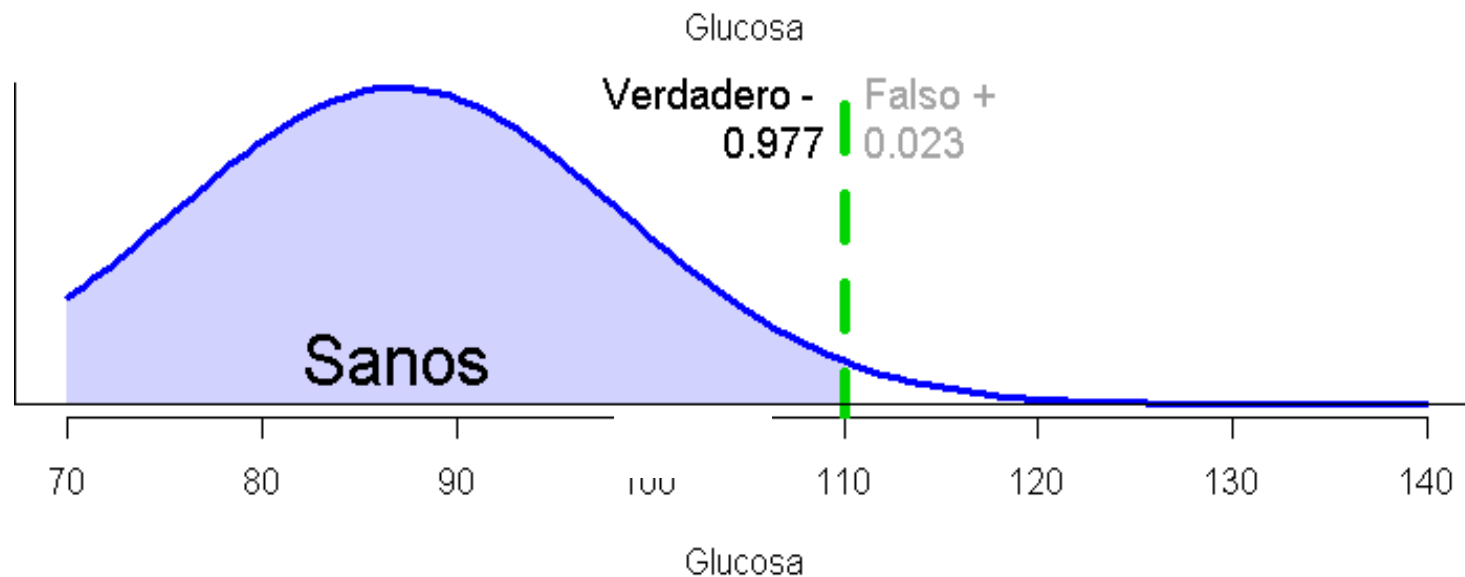
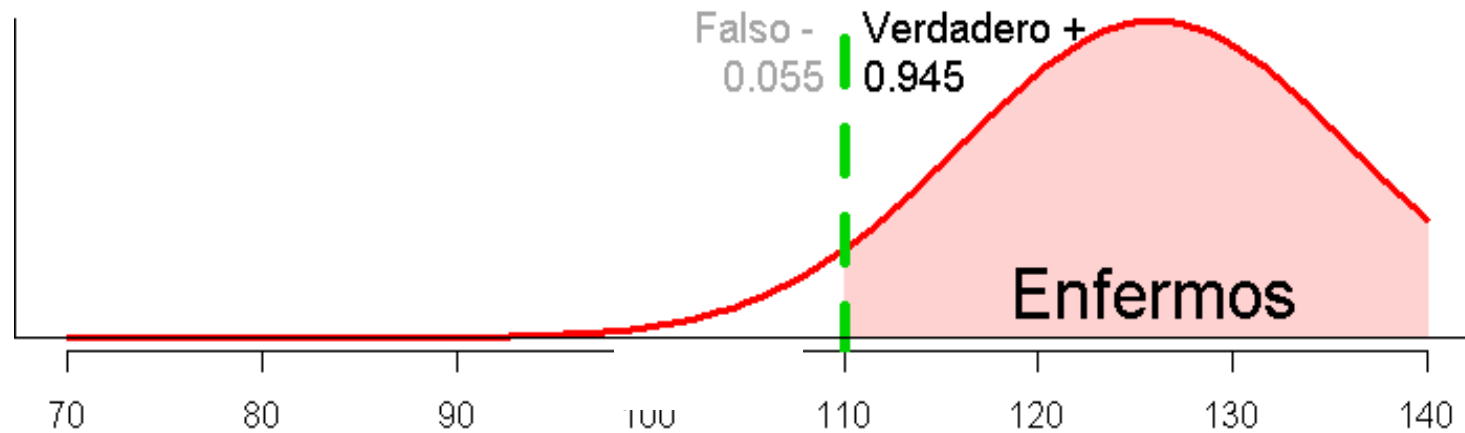
Medidas de exactitud diagnóstica: resultado numérico

Cuando la prueba diagnóstica arroja valores numéricos se debe definir que valor de corte se utilizará para clasificar a los individuos en la condición evaluada.



Acs anti citrulina anti-CCPhs (U/mL)	Diagnóstico	Estado de la Enfermedad
28	AR	Presente
5	AR	Presente
980	AR	Presente
1000	AR	Presente
121	AR	Presente
1000	AR	Presente
48	AR	Presente
8	AR	Presente
10	AR	Presente
1000	AR	Presente
803	AR	Presente
1000	AR	Presente
1000	AR	Presente
10	AR	Presente
78	AR	Presente
28	AR	Presente
1000	AR	Presente
14	ARJ	Ausente
5	ARJ	Ausente
1	ARJ	Ausente
13	FM	Ausente
9	FM	Ausente
16	FM	Ausente
8	FM	Ausente
14	LES	Ausente
9	LES	Ausente
14	EM	Ausente
10	Hipot. AI	Ausente
5	OA	Ausente
18	OA	Ausente
10	OA	Ausente
4	saludable	Ausente
10	saludable	Ausente
15	saludable	Ausente
4	saludable	Ausente

Buscando el valor de corte...



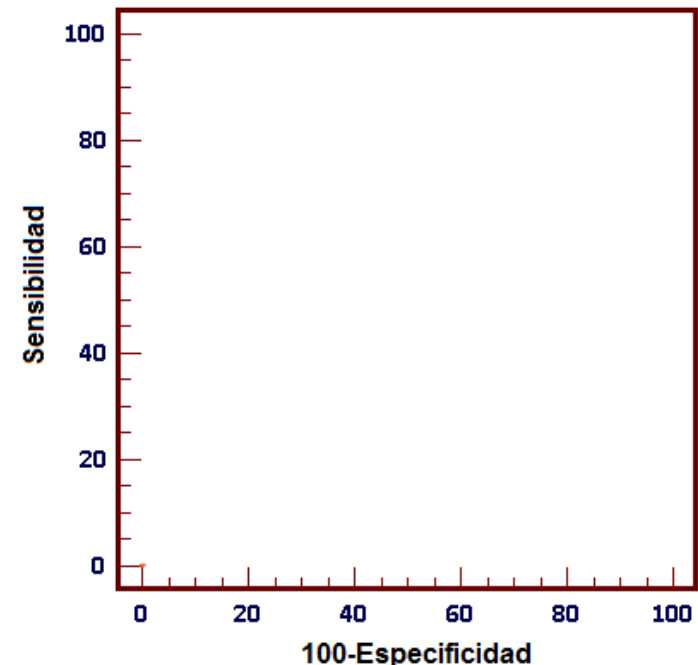
Medidas de exactitud diagnóstica: resultado numérico

Criterion	Sensitivity	95% CI	Specificity	95% CI
>=0	100,00	94,3 - 100,0	0,00	0,0 - 3,6
>0	100,00	94,3 - 100,0	1,94	0,3 - 6,9
>1	100,00	94,3 - 100,0	6,80	2,8 - 13,5
>2	96,87	89,1 - 99,5	11,65	6,2 - 19,5
>3	95,31	86,9 - 99,0	23,30	15,5 - 32,7
>4	89,06	78,7 - 95,5	32,04	23,2 - 42,0
>5	84,37	73,1 - 92,2	38,83	29,4 - 48,9
>6	82,81	71,3 - 91,1	43,69	33,9 - 53,8
>7	82,81	71,3 - 91,1	50,49	40,5 - 60,5
>8	79,69	67,8 - 88,7	54,37	44,3 - 64,2
>9	79,69	67,8 - 88,7	58,25	48,1 - 67,9
>10	75,00	62,6 - 85,0	66,99	57,0 - 75,9
>11	75,00	62,6 - 85,0	70,87	61,1 - 79,4
>12	75,00	62,6 - 85,0	71,84	62,1 - 80,3
>13	75,00	62,6 - 85,0	75,73	66,3 - 83,6
>14	75,00	62,6 - 85,0	82,52	73,8 - 89,3
>15	75,00	62,6 - 85,0	85,44	77,1 - 91,6
>16	75,00	62,6 - 85,0	87,38	79,4 - 93,1
>18	75,00	62,6 - 85,0	88,35	80,5 - 93,8
>19	73,44	60,9 - 83,7	88,35	80,5 - 93,8
>20	71,87	59,2 - 82,4	90,29	82,9 - 95,2
>22	71,87	59,2 - 82,4	91,26	84,1 - 95,9
>24	71,87	59,2 - 82,4	92,23	85,3 - 96,6
>25	70,31	57,6 - 81,1	92,23	85,3 - 96,6
>26	70,31	57,6 - 81,1	93,20	86,5 - 97,2

Se construyen tablas 2x2 con cada posible valor de corte y se calcula S y E.

Una manera más eficaz de mostrar la relación entre S y E es en forma gráfica. Se realiza un gráfico representando en el eje (y) **sensibilidad** y en el (x) se puede representar:

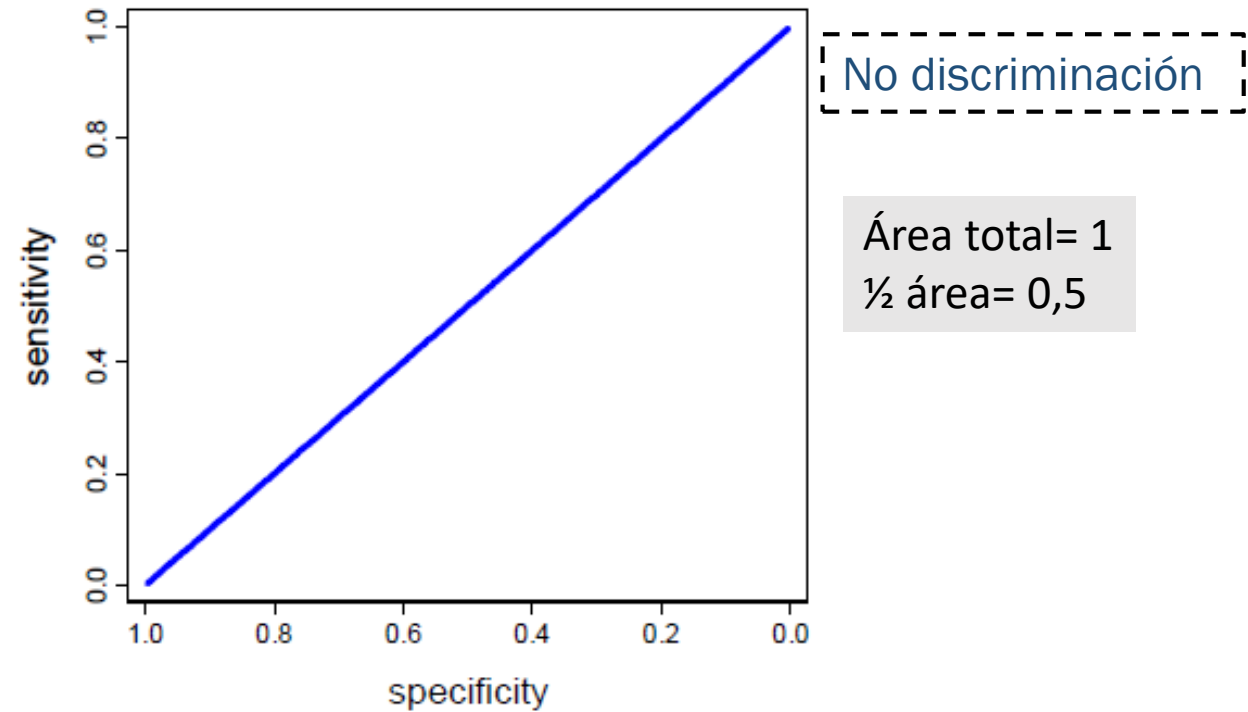
- **FP** o lo que es lo mismo (1 – E o 100-E)
- **Especificidad**



Curvas ROC

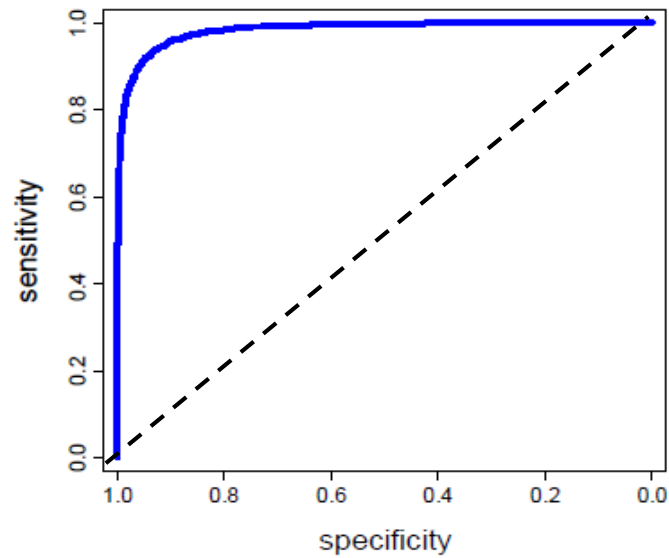
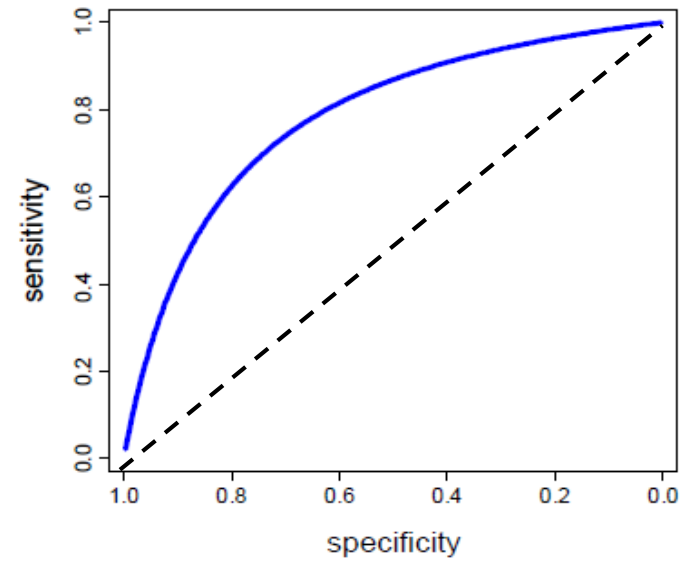
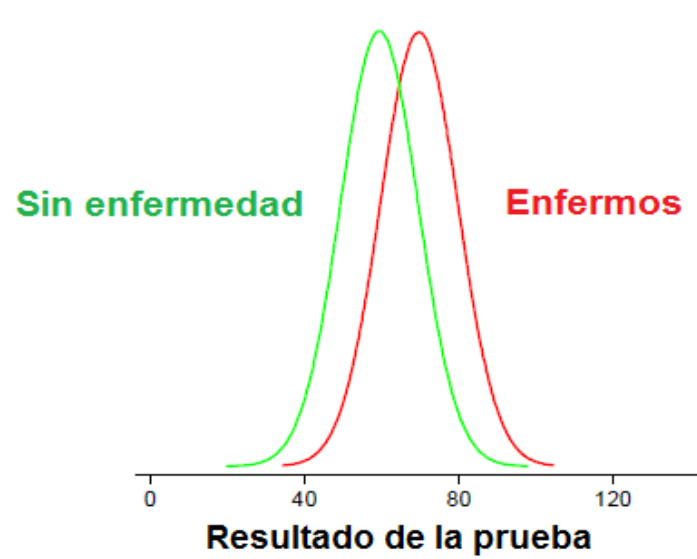
De la representación gráfica de todos los posibles valores de S y E se obtiene una curva, denominada **curva ROC** (curva receptor operador).

La posición de la curva ROC en el gráfico depende del grado de superposición de las distribuciones de los resultados de la prueba en enfermos y no enfermos.

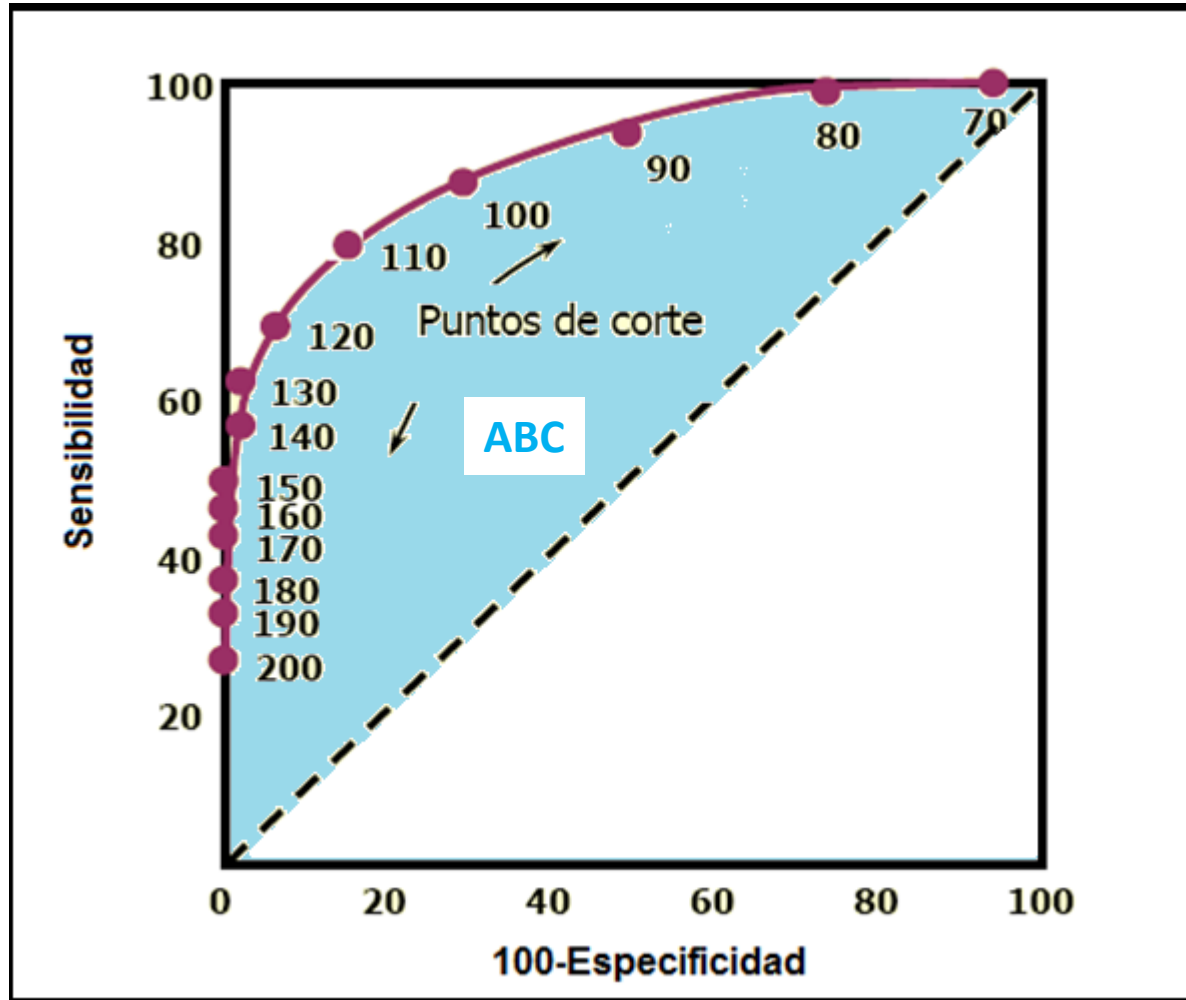


En esta situación la prueba da los mismos resultados para individuos “Enfermos” y “Sin enfermedad”

Curvas ROC



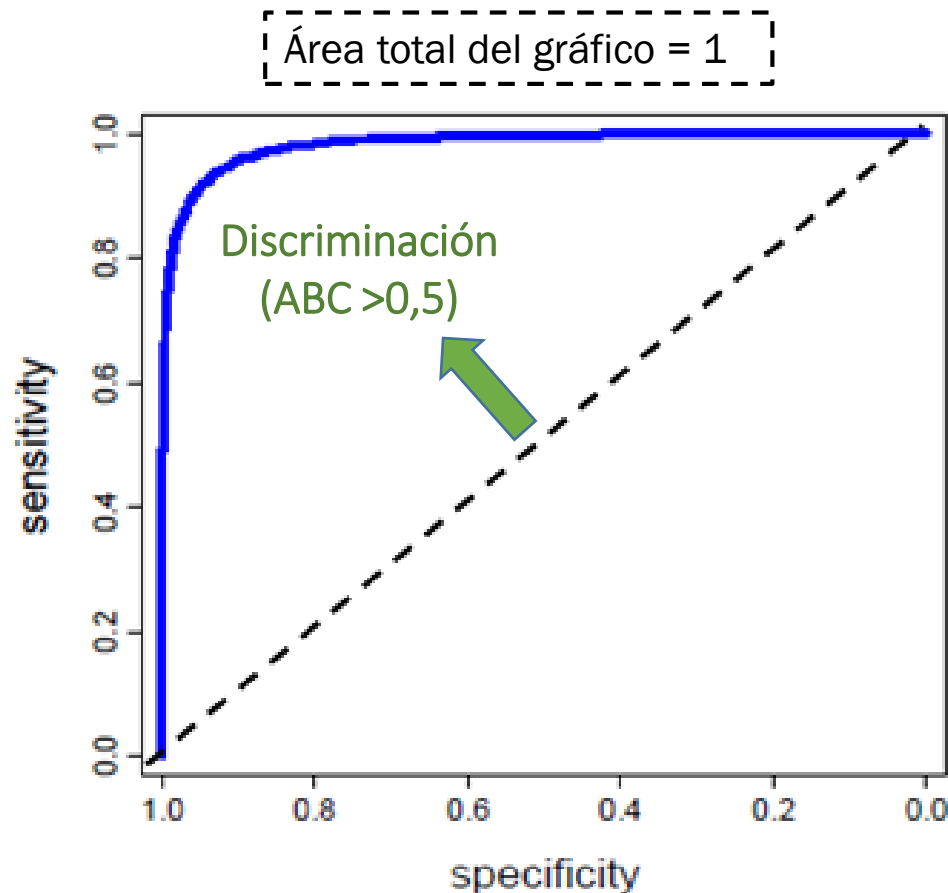
Curvas ROC



Con cada par de valores graficados se obtiene una curva; desde la diagonal punteada hasta el borde de la curva se delimita el **Área Bajo la Curva** (ABC o AUC en ingles).

El ABC es una medida de la exactitud diagnóstica global de la prueba; luego cada valor de corte tendrá su propia exactitud diagnóstica.

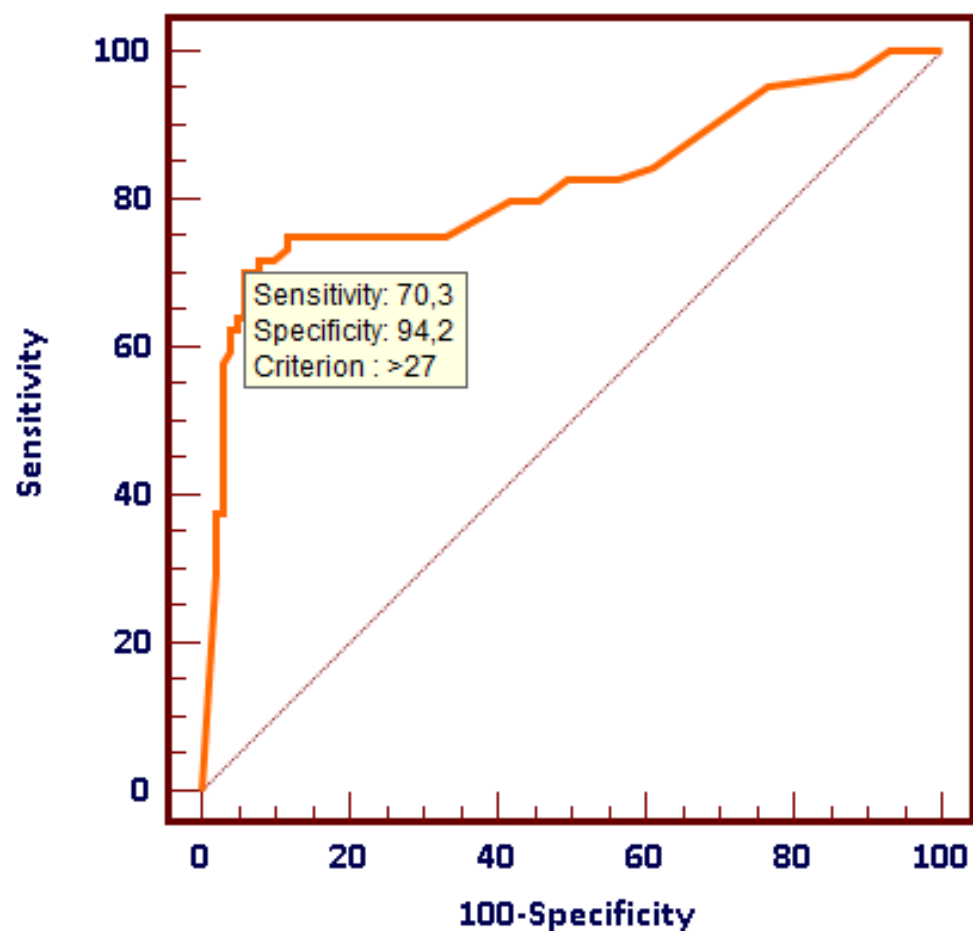
¿Cómo se interpreta el ABC de un curva ROC?



- El ABC adquiere valores entre 0,5 y 1
- Cuanto más cerca esté la curva de la esquina superior izquierda y cuanto mayor sea el ABC, mejor será la prueba para discriminar entre enfermos y no enfermos.
- Una prueba diagnóstica perfecta tiene un ABC= 1 y una no discriminadora un ABC= 0,5
- Para que el ABC sea estadísticamente significativa:
 - el IC95% no debe contener al 0,5
 - el $p < 0,05$

Para la elección del valor de corte se dispone de distintos métodos, el establecimiento del mismo dependerá del objetivo del estudio y/o utilidad clínica de la prueba (¿prueba de pesquisa? ¿prueba de confirmación diagnóstica?). Es decir, mayor sensibilidad o mayor especificidad según “uso previsto” de la prueba y utilidad clínica.

Ejemplo: valor de corte para Acs anti-Citrulina (anti-CCPhs)



Exactitud
diagnóstica
global
82,1%

Area under the ROC curve (AUC)	0,821
Standard Error	0,0355
95% Confidence Interval	0,754 to 0,876
z statistic	9,018
Significance level P (Area=0.5)	0,0001

Criterion	Sensitivity	95% CI	Specificity	95% CI	+LR	-LR
>=0	100,00	94,3 - 100,0	0,00	0,0 - 3,6	1,00	
>0	100,00	94,3 - 100,0	1,94	0,3 - 6,9	1,02	0,00
>1	100,00	94,3 - 100,0	6,80	2,8 - 13,5	1,07	0,00
>2	96,87	89,1 - 99,5	11,65	6,2 - 19,5	1,10	0,27
>3	95,31	86,9 - 99,0	23,30	15,5 - 32,7	1,24	0,20
>4	89,06	78,7 - 95,5	32,04	23,2 - 42,0	1,31	0,34
>5	84,37	73,1 - 92,2	38,83	29,4 - 48,9	1,38	0,40
>6	82,81	71,3 - 91,1	43,69	33,9 - 53,8	1,47	0,39
>7	82,81	71,3 - 91,1	50,49	40,5 - 60,5	1,67	0,34
>8	79,69	67,8 - 88,7	54,37	44,3 - 64,2	1,75	0,37
>9	79,69	67,8 - 88,7	58,25	48,1 - 67,9	1,91	0,35
>10	75,00	62,6 - 85,0	66,99	57,0 - 75,9	2,27	0,37
>11	75,00	62,6 - 85,0	70,87	61,1 - 79,4	2,58	0,35
>12	75,00	62,6 - 85,0	71,84	62,1 - 80,3	2,66	0,35
>13	75,00	62,6 - 85,0	75,73	66,3 - 83,6	3,09	0,33
>14	75,00	62,6 - 85,0	82,52	73,8 - 89,3	4,29	0,30
>15	75,00	62,6 - 85,0	85,44	77,1 - 91,6	5,15	0,29
>16	75,00	62,6 - 85,0	87,38	79,4 - 93,1	5,94	0,29
>18	75,00	62,6 - 85,0	88,35	80,5 - 93,8	6,44	0,28
>19	73,44	60,9 - 83,7	88,35	80,5 - 93,8	6,30	0,30
>20	71,87	59,2 - 82,4	90,29	82,9 - 95,2	7,40	0,31
>22	71,87	59,2 - 82,4	91,26	84,1 - 95,9	8,23	0,31
>24	71,87	59,2 - 82,4	92,23	85,3 - 96,6	9,25	0,30
>25	70,31	57,6 - 81,1	92,23	85,3 - 96,6	9,05	0,32
>26	70,31	57,6 - 81,1	93,20	86,5 - 97,2	10,35	0,32
>27 *	70,31	57,6 - 81,1	94,17	87,7 - 97,8	12,07	0,32
>28	65,62	52,7 - 77,0	94,17	87,7 - 97,8	11,27	0,37

AR (n= 64)

No AR (n= 103)

Ejemplo

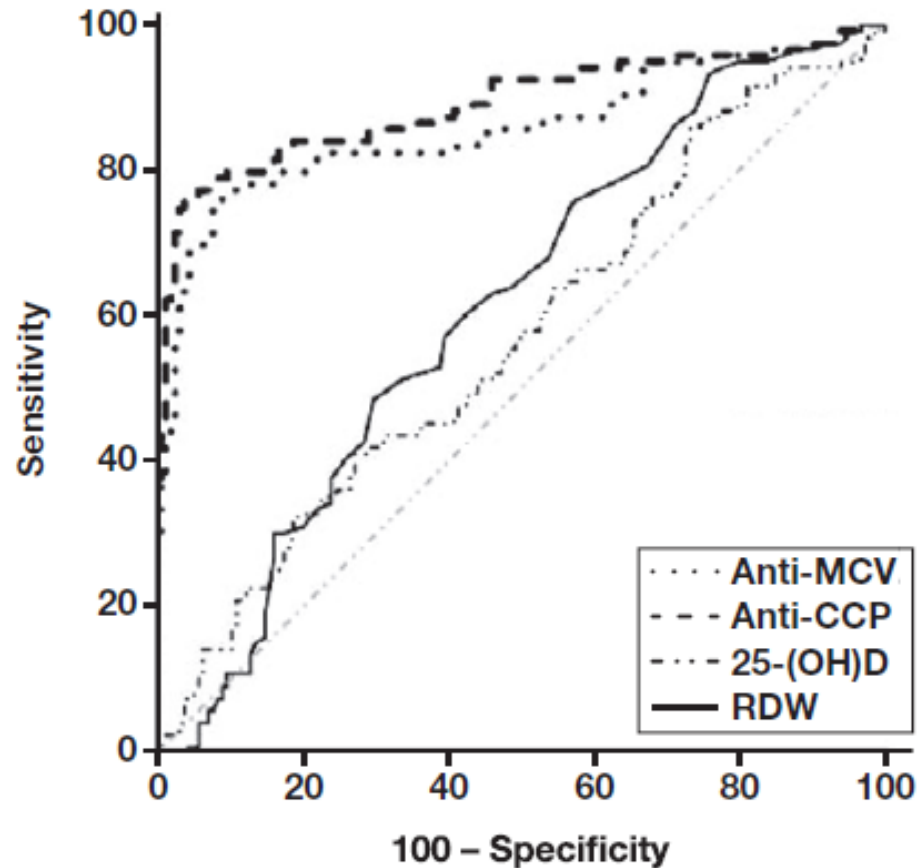


Figure 1

Specificity and sensitivity (per 100) of anti-mutated citrullinated vimentin (anti-MCV), anti-citrullinated peptide (anti-CCP), 25-hydroxyvitamin D (25-[OH]D), and red blood cell distribution width (RDW).

Table 3. ROC Curve Analysis Results^a

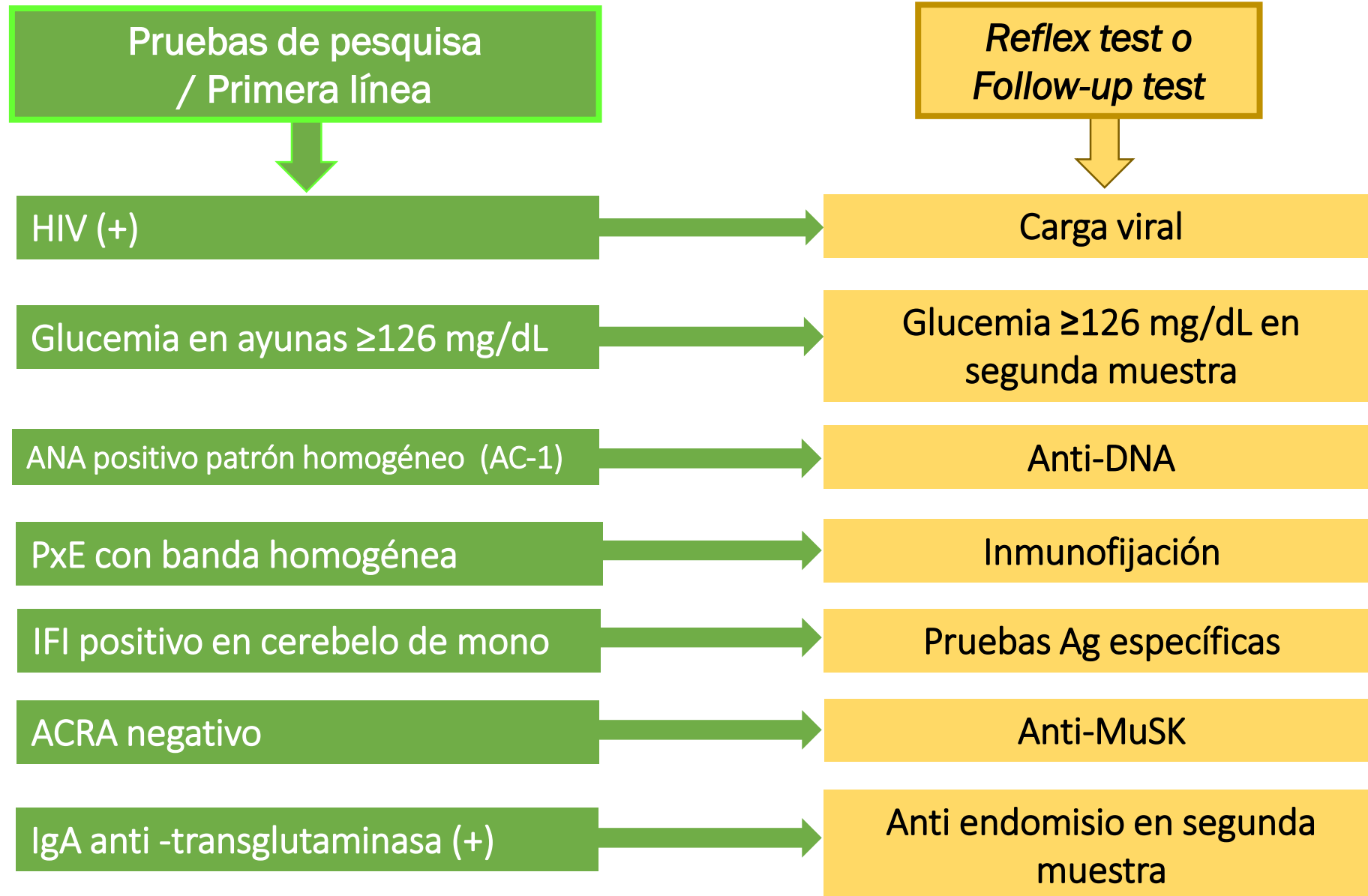
Analyte	AUC	Cutoff Value	P Value (95% CI)
Anti-MCV	0.857	55.32	<.001 (.81-.91)
Anti-CCP	0.890	33.40	<.001 (.85-.93)
RDW	0.611	13.85	.002 (.54-.68)
25-(OH)D	0.569	30.84	.049 (0.50-.64)

ROC, receiver operating characteristic; AUC, area under the curve; CI, confidence interval; anti-MCV, anti-mutated citrullinated vimentin; anti-CCP, anti-citrullinated peptide; RDW, red blood cell distribution width; 25-(OH)D, 25-hydroxyvitamin D; RA, rheumatoid arthritis.

^aIn total, 119 patients with RA were considered to be the disease group. The 114 patients without RA, combined with the 40 healthy people, were considered to be the no-disease group.

Se puede representar mas de una curva ROC por gráfico y se dispone de test estadísticos para evaluar si existen diferencias estadísticamente significativas entre las ABC.

Pruebas de primera y segunda línea

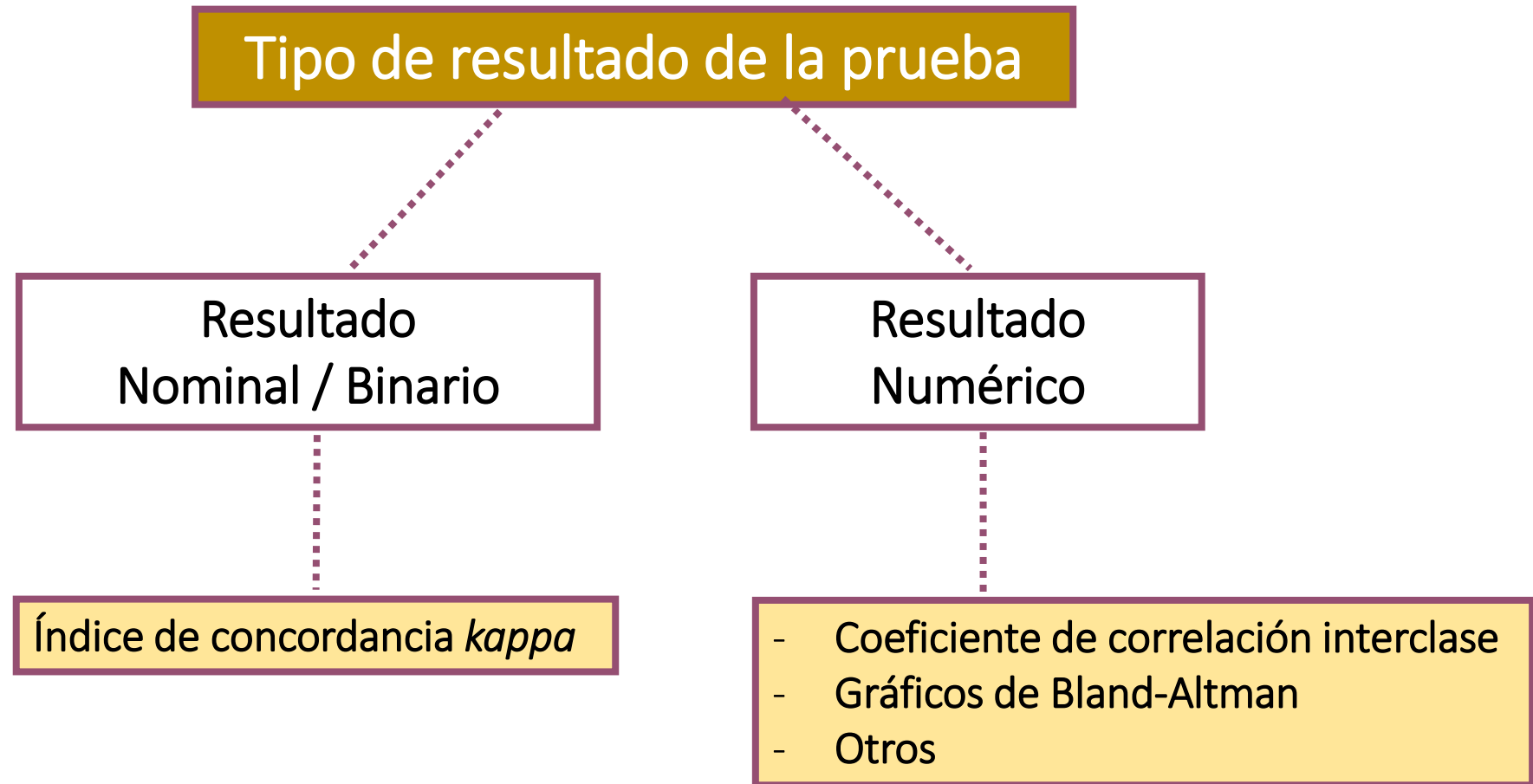


¿Qué sucede si el interés es establecer si hay acuerdo o concordancia en el resultado dado por dos pruebas?

El término **concordancia** se deriva de la expresión latina *concordare*, cuyo significado hace referencia a que hay 'correspondencia o conformidad' de una cosa con otra.



Concordancia según el tipo de variable



Resultado nominal / binario: coeficiente de concordancia *kappa*

Prueba A / observador A

	Positivo	Negativo
Positivo	Acuerdo	Desacuerdo
Negativo	Desacuerdo	Acuerdo

Prueba B /
Observador B

Prueba B /
Observador B

Prueba A / Observador A

	+	++	+++	++++
+	Acuerdo	Desacuerdo	Desacuerdo	Desacuerdo
++	Desacuerdo	Acuerdo	Desacuerdo	Desacuerdo
+++	Desacuerdo	Desacuerdo	Acuerdo	Desacuerdo
++++	Desacuerdo	Desacuerdo	Desacuerdo	Acuerdo

El coeficiente de concordancia **Kappa** (*k*) es una medida más robusta que el simple porcentaje de concordancia (“concordancia absoluta”), ya que para su cálculo ajusta el efecto del azar en la proporción de la concordancia observada. Este ajuste se debe a que parte de los “acuerdos” observados pueden deberse al azar. Suele expresarse junto al IC95%. El coeficiente **Kappa ponderado** (o *weighted kappa*) es una medida estadística que evalúa la concordancia entre dos evaluadores cuando los resultados se expresan en **escala ordinal** (ej.: cruces).

Kappa de Cohen mide el acuerdo en tabla 2x2, mientras que **kappa de Fleiss** se emplea en tablas con más de 2 filas y 2 columnas.

Coeficiente de concordancia Kappa	Interpretación de la concordancia
< 0	Menor que el azar
0	Igual al azar
0,10 a 0,30	Mala
0,31 a 0,50	Regular
0,51 a 0,71	Buena
0,71 a 0,90	Muy buena
0,91 a 1	Excelente

Ejemplo

Pacientes con AR (n= 58)

ELISA anti-CCP

ELISA anti-CCPhs

	Positivo	Negativo
Positivo	40	1
Negativo	5	12

Concordancia absoluta= $(40+12/58)*100= 89,7\%$

Inter-rater agreement

Observer A:

Observer B:

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
Class 1	40	1				
Class 2	5	12				
Class 3						
Class 4						
Class 5						
Class 6						

☐ Weighted Kappa ☒ Linear weights ☐ Quadratic weights

Kappa	0,732
Standard error	0,104
95% CI	0,529 to 0,935

<https://www.medcalc.org/>

Ejemplo

Patrones en HEp-2 por IFI (n= 56)

(<https://www.anapatterns.org/trees-2021.php>)

Observador B

		Observador A			
		AC-0	AC-1	AC-2	AC-4
AC-0		25	0	0	0
AC-1		0	10	3	0
AC-2		0	4	4	0
AC-4		0	0	0	10

Concordancia absoluta= $(25+10+4+10/56)*100= 87,5\%$

The screenshot shows a software window titled "Inter-rater agreement". It contains a table for comparing two observers, A and B, across six classes. The table is a 6x6 grid where the columns represent Observer A's classes (Class 1 to Class 6) and the rows represent Observer B's classes (Class 1 to Class 6). The values in the cells represent the number of cases for each combination. Below the table, there are options for "Weighted Kappa" (unchecked) and "Linear weights" (selected). At the bottom, the calculated Kappa value is 0,820, the Standard error is 0,064, and the 95% CI is 0,695 to 0,945. There are buttons for "Help", "Test", and "Exit".

		Observer A:					
Observer B:		Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
Class 1		25	0	0	0		
Class 2		0	10	3	0		
Class 3		0	4	4	0		
Class 4		0	0	0	10		
Class 5							
Class 6							

☐ Weighted Kappa ☒ Linear weights
☐ Quadratic weights

Kappa 0,820
Standard error 0,064
95% CI 0,695 to 0,945

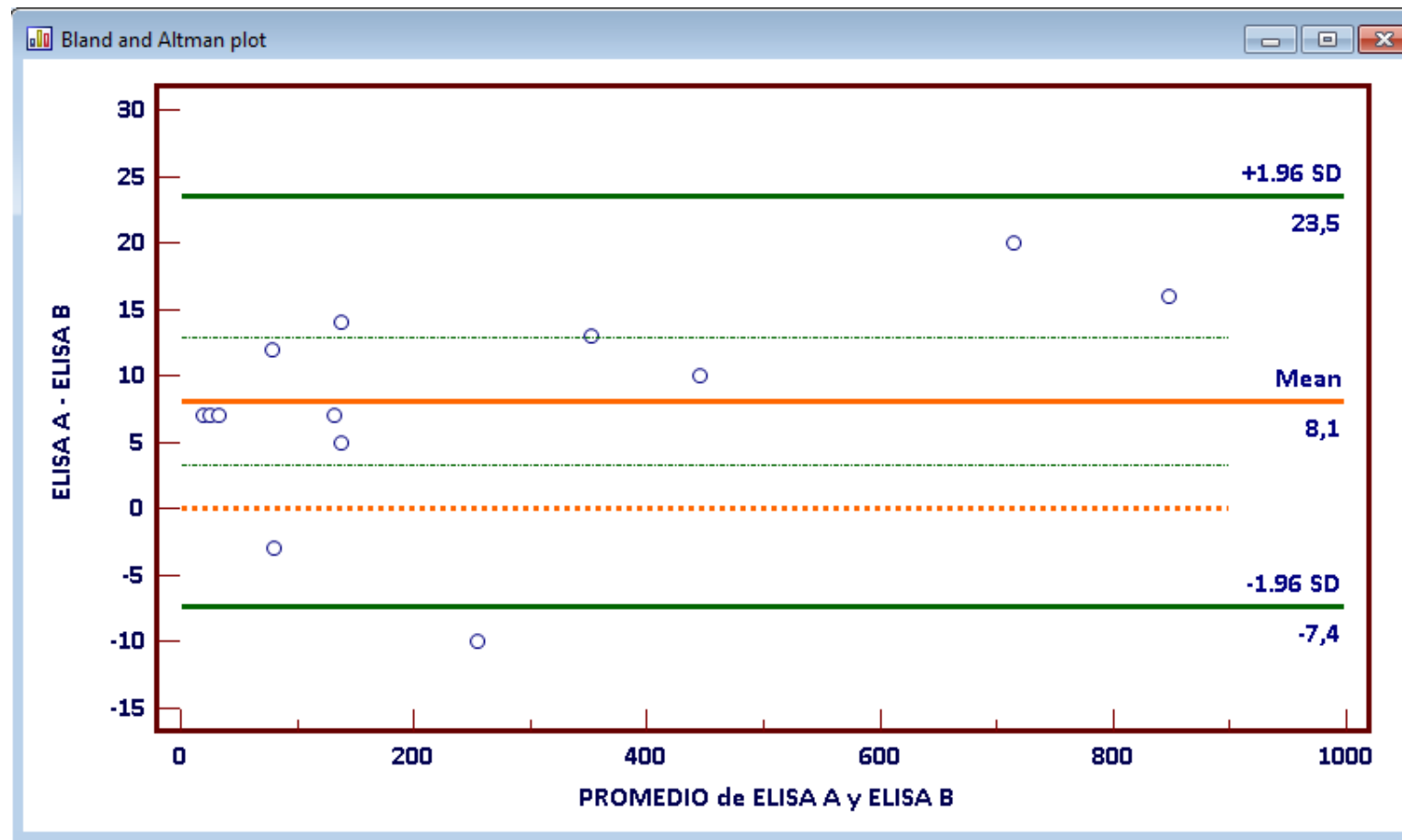
Help Test Exit

<https://www.medcalc.org/>

Es importante establecer la concordancia absoluta esperada y el coeficiente *kappa* “aceptable” antes de realizar la comparación, como así también ponderar las diferencia, cuales desacuerdos serían aceptados y cuales no (ej.: no se acepta que el observador A obtenga un resultado “AC-0” y el observador B un “AC-1”).

Resultado numérico: gráficos de Bland-Altman

Métodos a evaluar (U/mL)	
ELISA "A"	ELISA "B"
144	130
450	440
22	15
725	705
85	73
135	128
140	135
250	260
29	22
79	82
358	345
856	840
36	29



<https://www.medcalc.org/>

NO son medidas de concordancia:

- Correlación lineal (coeficiente de correlación de *Pearson* o *Spearman*)
- Regresión lineal

BMJ Open STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration

Jérémie F Cohen,^{1,2} Daniël A Korevaar,¹ Douglas G Altman,³ David E Bruns,⁴ Constantine A Gatsonis,⁵ Lotty Hooft,⁶ Les Irwig,⁷ Deborah Levine,^{8,9} Johannes B Reitsma,¹⁰ Henrica C W de Vet,¹¹ Patrick M M Bossuyt¹

- La iniciativa STARD (*Standards for Reporting of Diagnostic Accuracy Studies*) tiene el objetivo de armonizar las publicaciones de estudios sobre exactitud diagnóstica con el fin de brindar información completa y transparente.
- STARD 2015 es un *up date* de la guía publicada en 2003 (STARD 2003).
- STARD contiene una lista de verificación con 30 ítems que deben ser informados en cualquier estudio sobre exactitud diagnóstica.
- Además, tiene como objetivo brindar toda la información necesaria para facilitar la aplicación de QUADAS-2, una herramienta que evalúa la calidad de las publicaciones sobre exactitud diagnóstica.

¿Quien debe realizar los estudios de exactitud diagnóstica?

Pregunta	Respuesta
¿Quién debe realizar los estudios de exactitud diagnóstica?	El fabricante, según IVDR, FDA, ISO 15189.
¿Qué debe hacer el laboratorio (usuario)?	Verificar desempeño local (precisión, sesgo, linealidad, interferencias).
¿Cómo se evalúa la calidad de un estudio de exactitud diagnóstica?	STARD 2015, QUADAS-2, GRADE, y requerimientos regulatorios (IVDR, FDA).

De acuerdo con la norma ISO 15189, el reglamento CE IVDR, las directivas CAP y el nuevo reglamento IVD de 2017, el fabricante es responsable de la validación clínica de una prueba con marcado CE/FDA que el laboratorio utilice sin modificaciones.

Consideraciones finales

La exactitud diagnóstica de una prueba es evaluada por la capacidad de la prueba de correctamente clasificar a un individuo dentro de dos categorías posibles.

Para establecer sensibilidad y especificidad diagnóstica se debe realizar la clasificación de los individuos en forma “inequívoca”, con el “método de referencia” o la mejor evidencia disponible para el diagnóstico de la enfermedad o condición de interés.

La estadística y el análisis de los datos dependerá del tipo de resultado de la prueba a evaluar. Todos los estadísticos calculados deben estar acompañados por sus respectivos intervalos de confianza (IC95%).

Los *likelihood ratio*, también referidos como razón de verosimilitud, razón de probabilidad o cociente de probabilidad, son una de las mejores herramientas de exactitud diagnóstica para evaluar la utilidad clínica de una prueba.

Es de suma importancia el informe de los resultados de pruebas diagnósticas, teniendo en cuenta: la “intención de uso” o “uso previsto” de la prueba, la utilidad clínica y la seguridad del paciente.

Para la elaboración y reporte de los resultados de estudios sobre exactitud diagnóstica se debe adherir a las recomendaciones publicadas en las guías vigentes que abordan el tema.



MUCHAS GRACIAS POR SU ATENCIÓN