

Pruebas Diagnósticas en Epidemiología, Sensibilidad, Especificidad, Valores Predictivos, Concordancia

Prof. Alfonso J. Rodríguez-Morales,
MD, MSc, DTM&H, FRSTMH(Lon), FFTM RPCS(Glasg), PhD(c)
Médico tropicalista y salubrista - Investigador

Docente Transitorio, Coordinador, Factores de Riesgo, **Departamento de Medicina Comunitaria,**
Facultad de Ciencias de la Salud (**Centro Adherido de la Red Cochrane Iberoamericana para Colombia**),
Universidad Tecnológica de Pereira (UTP), Pereira, Risaralda, Colombia.

Instructor de Talleres de MBE en Venezuela, Perú y Brasil.

Sentinel Reader in **MORE (McMaster Online Rating of Evidence)** de **Evidence-Based Journals Group,**
McMaster University, Hamilton, Canada.

Editor, **Journal of Infection in Developing Countries (JIDC).**

Editorial Board, **Travel Medicine & Infectious Diseases (TMAID)**

Editor Asistente, **Revista Médica de Risaralda (RMR).**

E-mail: arodriguezm@utp.edu.co



Participants and Sample

Participants

Research is a formalised curiosity. It is poking and prying with a reason.

Zora Neal Hurston (African-American novelist, 1891–1960, www.bartelby.com)

Sample size

It is most important to have a beautiful theory. And if the observations don't support it, don't be too distressed, but wait a bit and see if some error in the observations doesn't show up.

Paul Dirac (theoretical physicist, 1980)

Peat J, Elliott E, Baur L, Keena V.
Scientific Writing – Easy when you
know how. BMJ Books, London, 2002.

	Good	Bad
Accept	OK!	beta
Reject	alpha	OK!

Courtesy Cayman Systems

Diagnóstico

- La medicina es una ciencia de probabilidades y un arte de manejar la incertidumbre.
- El diagnóstico siempre se enfrenta con información insuficiente.
- Los instrumentos diagnósticos se consideran medios para reducir la incertidumbre.
- Todavía son potentes instrumentos diagnósticos, la historia clínica y el examen físico.

Umbral de la prueba (Pauker y Kassirer, 1980)

Límite de descarte

Límite de diagnóstico



No tiene la
Enfermedad

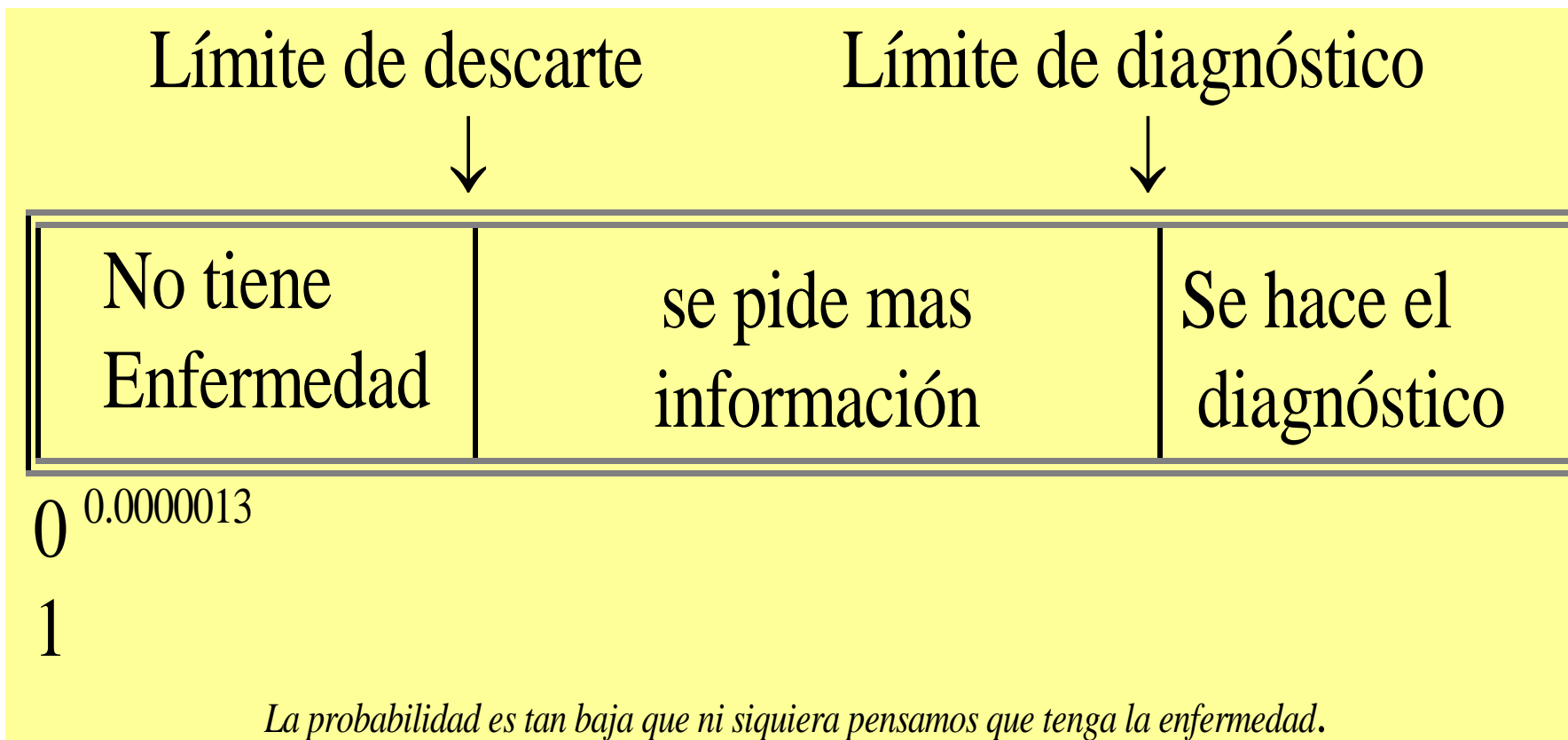
se pide mas
información

Se hace el
diagnóstico

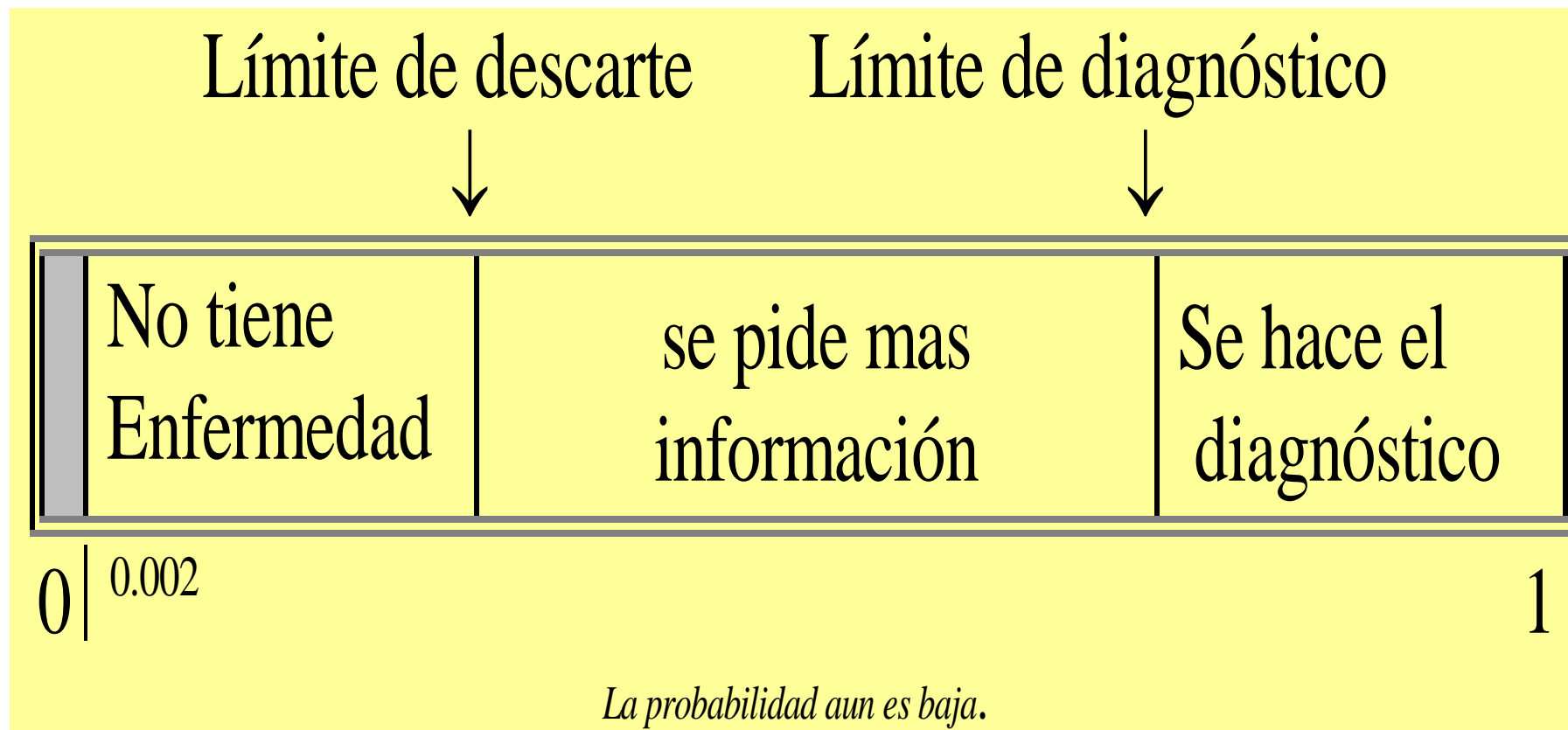
0

1

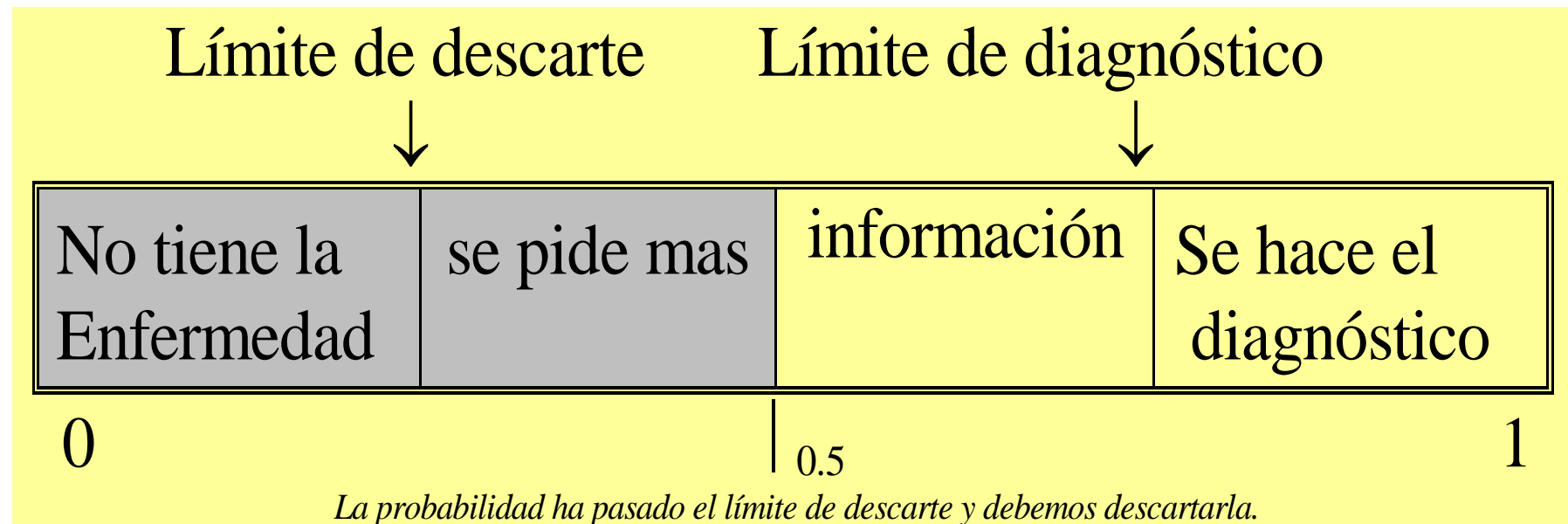
Un paciente de la población general
incidencia 50/100 000 , $P=0.0000013$



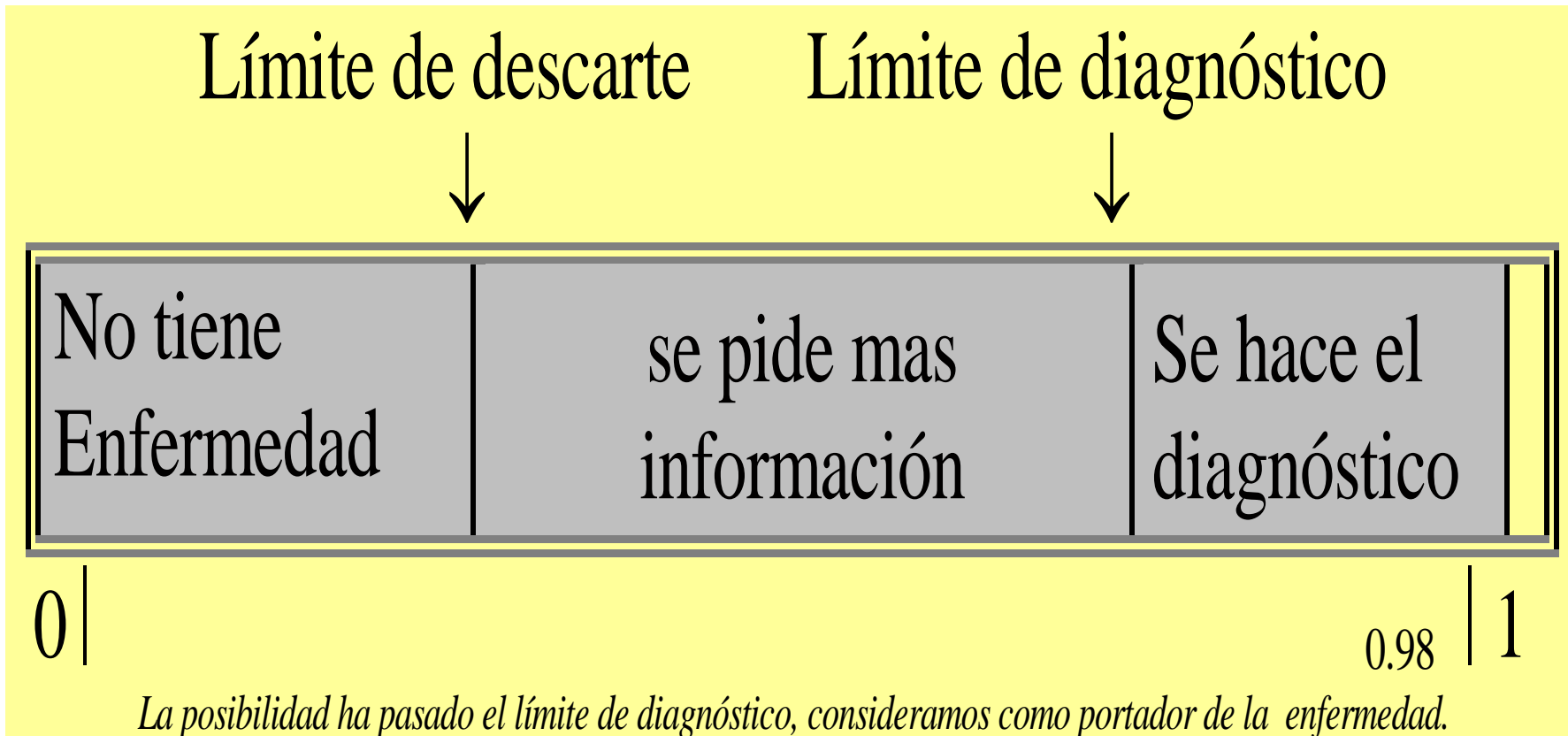
Paciente que acude a la emergencia, donde $1/500$ acuden por Infarto de miocardio , $P= 0.002$



Paciente varón de 50 años con precordialgia
intensa de mas de 30 minutos, pálido, sudoroso,
sensación de que se va morir, $P= 0.5$



El mismo paciente tiene $CPK = 400$
Para $CPK > 280$ $P = 0.98$.

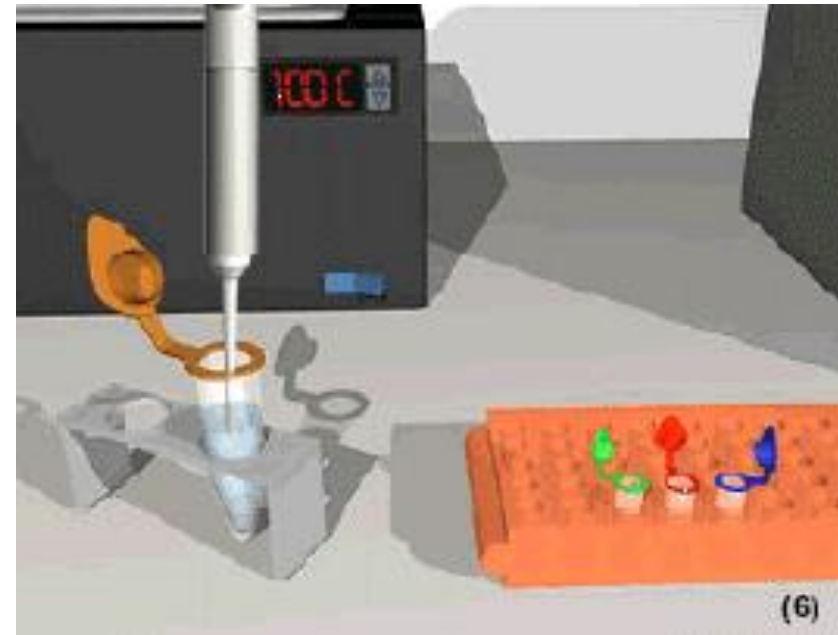


¿Cómo disminuir la incertidumbre?

- Con mas información
- Por ejemplo:
 - Mas datos de historia clínica
 - pruebas auxiliares

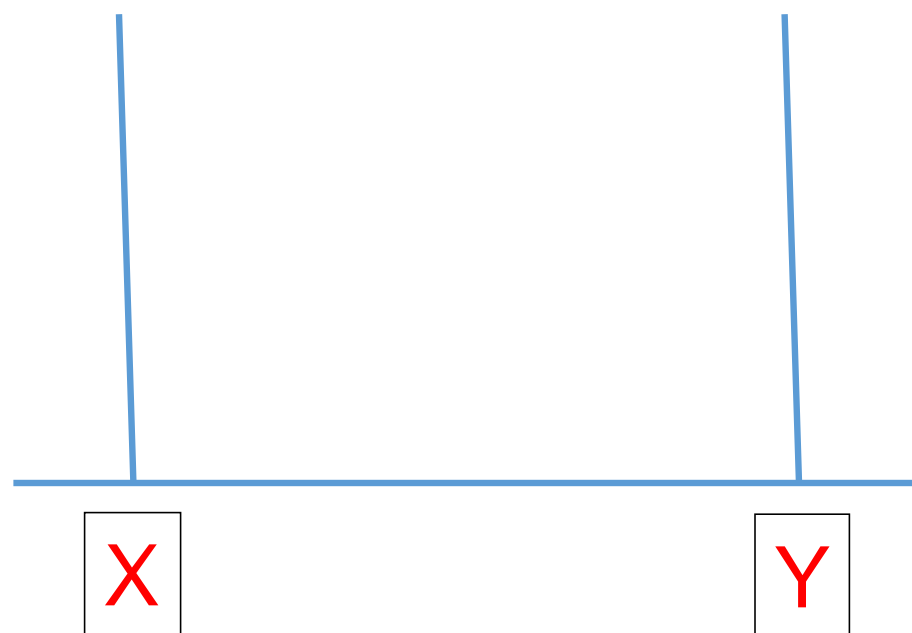
Diagnóstico

- Hoy se dispone de un gran volumen tecnología auxiliar de diagnóstico.
- Los principios fundamentales de las pruebas diagnósticas deben ayudar a identificar que pruebas van a aumentar la precisión diagnóstica y cuales sólo van a incrementar el costo
- El médico debe aprender cuando asumir riesgos para aumentar la certeza y cuando tolerar la incertidumbre



Condiciones de una prueba diagnóstica perfecta

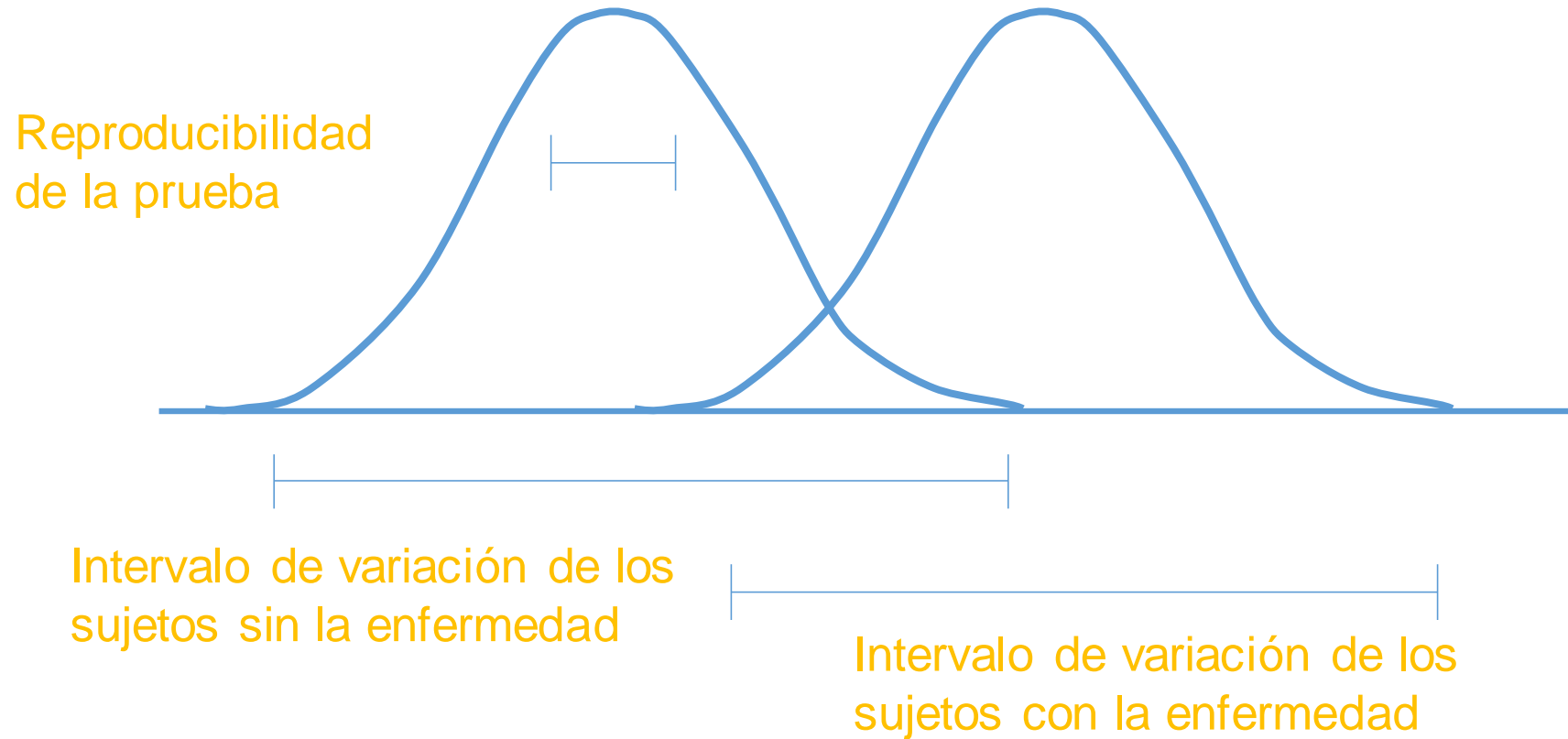
1. Todos los individuos sin la enfermedad tienen un valor uniforme en la prueba
2. Todos los individuos con la enfermedad tienen un valor uniforme, pero distinto en la prueba.
3. Todos los resultados de las pruebas son consistentes con los resultados del grupo de los enfermos y del de los sanos



X = valor para el sujeto
con la enfermedad

Y = valor para el sujeto
sin la enfermedad

Tipos de variaciones que afectan las pruebas diagnósticas



Marco de la prueba de una prueba

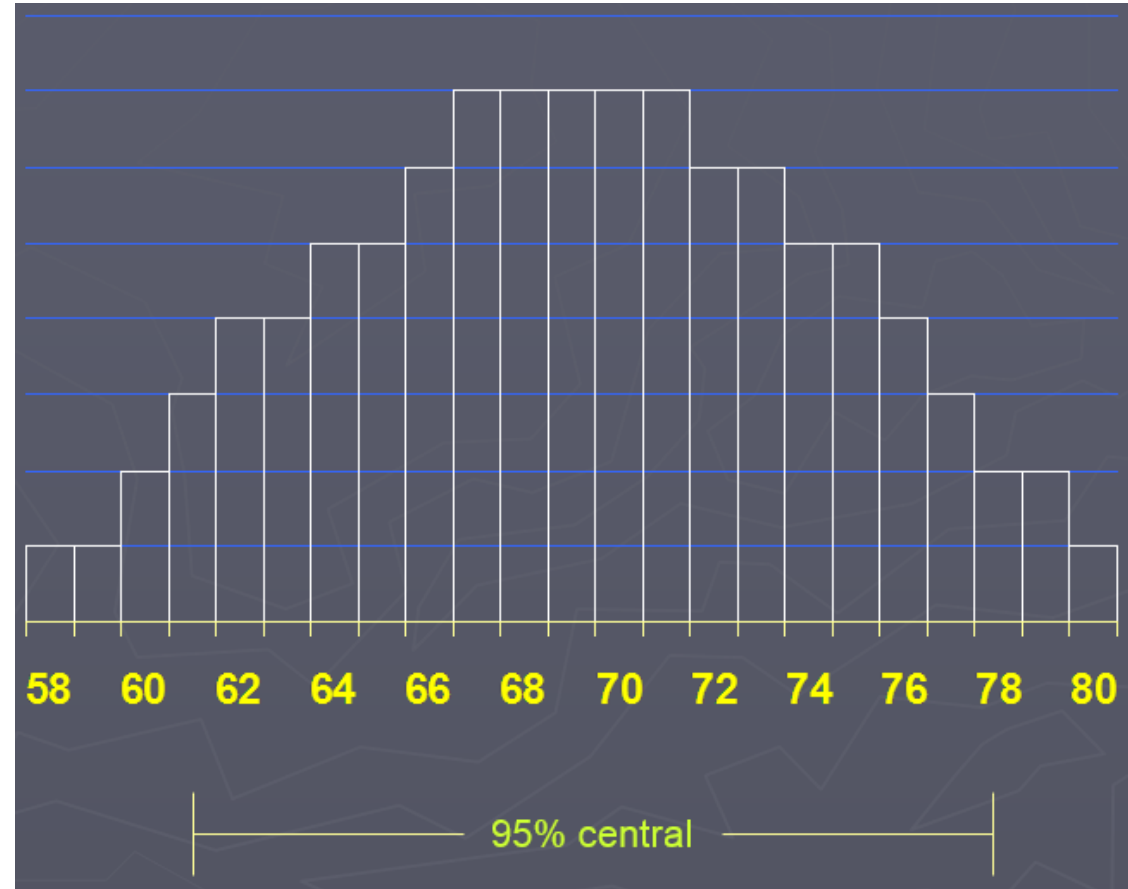


El Intervalo de normalidad

- ▶ Es un esfuerzo por cuantificar un conjunto de valores que existen en individuos considerados sanos. (Presión arterial, tamaño del hígado, niveles de glucosa en sangre, etc.)
- ▶ El intervalo excluye a 5% considerados sanos. Podría variar por el azar, por cambios fisiológicos o patológicos secundarios o no asociados con la enfermedad.
- ▶ Es descriptivo, no es diagnóstico.

El Intervalo de normalidad

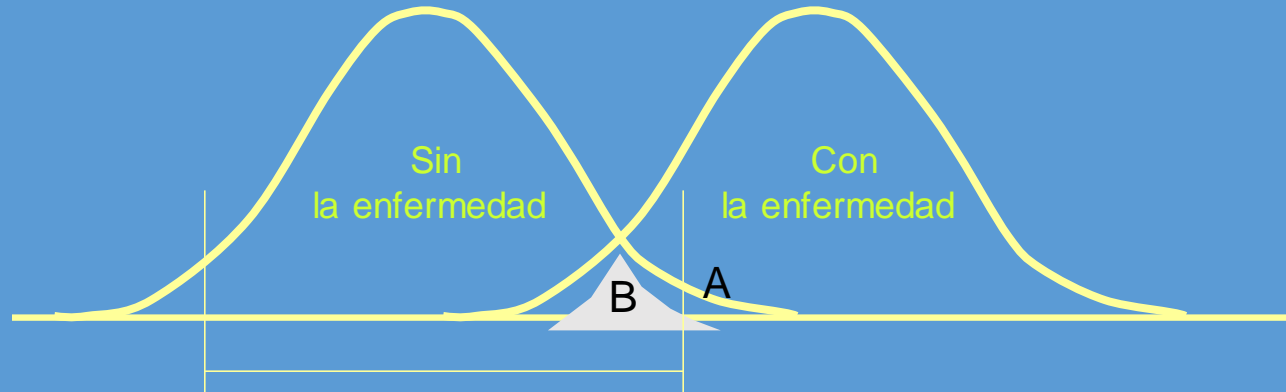
1. Identificar un grupo de referencia.
2. Ejecutar la prueba de interés en todos los individuos.
3. Representar gráficamente sus resultados
4. Calcular el intervalo de lo normal. La media ± 2 desviaciones estándar



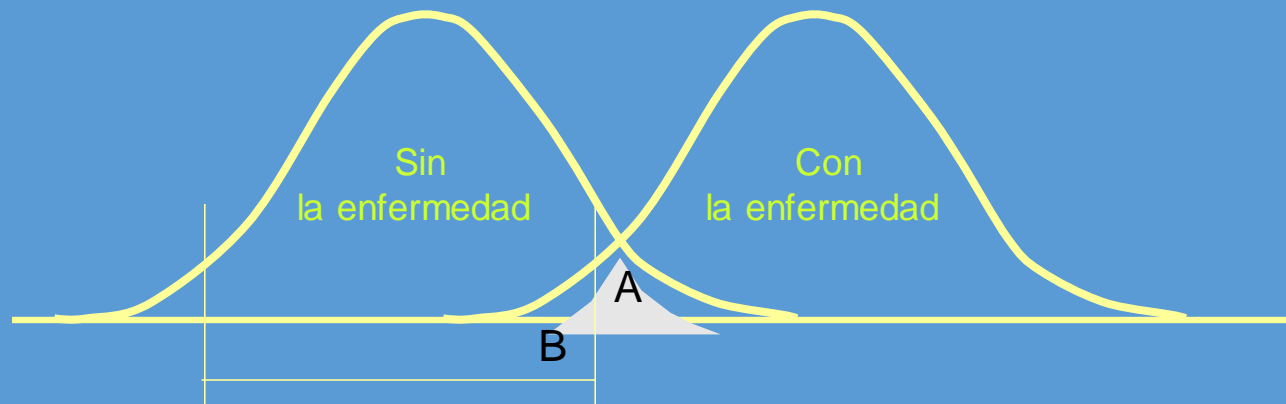
El Intervalo de normalidad: Principios básicos

1. 5% de los "sanos" estarán fuera del intervalo. "anormal" no es sinónimo de "enfermo".
2. Los valores dentro del intervalo no garantizan ausencia de enfermedad.
3. Los cambios incluidos dentro del intervalo podrían ser patológicos.
4. El intervalo se construye empleando un grupo concreto de personas.
5. El intervalo de la normalidad no se debe confundir con el intervalo de lo deseable.
6. Los límites del intervalo pueden modificarse con fines diagnósticos

El Intervalo de normalidad: falsos positivos, falsos negativos



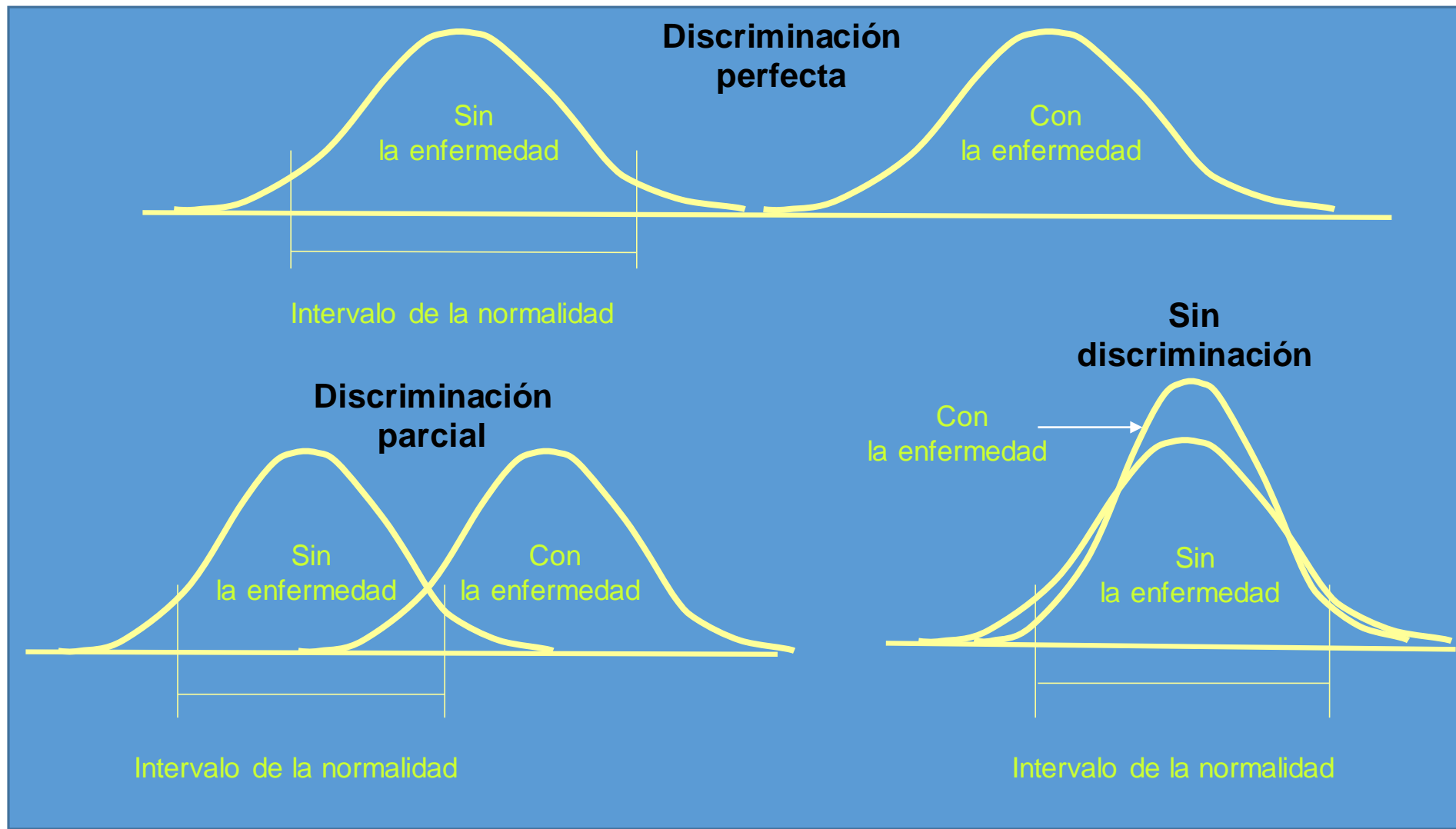
Intervalo de la normalidad N° 1



Intervalo de la normalidad N° 2



El Intervalo de normalidad: Discriminación diagnóstica



Definición de la Enfermedad: La “prueba de oro”



- ▶ La prueba o criterio para definir “inequívocamente” una enfermedad. Prueba más antigua o más aceptada.
- ▶ Se parte del supuesto que la “prueba de oro” logrará en el 100% de los casos un diagnóstico correcto.
- ▶ Requisito para examinar la utilidad diagnóstica de una prueba.
- ▶ La nueva prueba sólo puede tratar de alcanzar ser tan óptima como la prueba de oro.
- ▶ El investigador debe cerciorarse de que la prueba de oro utilizada ha sido establecido como la mejor referencia posible

Discriminación diagnóstica: Sensibilidad y Especificidad

- ▶ Son medidas del valor del diagnóstico
- ▶ Miden la discriminación de la nueva prueba comparada con la prueba de oro, que por definición tiene una sensibilidad y especificidad de 100%.
- ▶ Son características intrínsecas de una prueba sea que la enfermedad es frecuente o rara.
- ▶ Permiten comparar el rendimiento de una prueba con otra.
- ▶ La sensibilidad: Proporción de enfermos identificados correctamente por la prueba.
- ▶ La especificidad: Proporción de sanos identificados correctamente por la prueba.



Discriminación diagnóstica: Sensibilidad y Especificidad

1. Seleccionar una prueba de oro.
2. Seleccionar un grupo que según la prueba de oro padecen la enfermedad y otro que según la misma, están sanos.
3. Aplicar la prueba en investigación a todos los individuos.

Prueba de Estudio

Prueba de oro Enfermos

Prueba de oro Sanos

+

-

a = Individuos enfermos y positivos	b = Individuos sanos y positivos
c = Individuos enfermos y negativos	d = Individuos sanos y negativos

$a + c =$ Total de Individuos enfermos

$b + d =$ Total de Individuos sanos

$$\text{Sensibilidad} = a / a + c$$

$$\text{Especificidad} = d / b + d$$

Discriminación diagnóstica: Sensibilidad y Especificidad

	Prueba de Estudio	Prueba de oro Enfermos	Prueba de oro Sanos
Sensibilidad = $a / a + c$			
$400 / 500 = 0,8$ $= 80\%$	+	400	50
Especificidad = $d / b + d$	-	100	450
$450 / 500 = 0,9$ $= 90\%$		500	500



Discriminación diagnóstica: Sensibilidad y Especificidad

Prevalencia de 50%

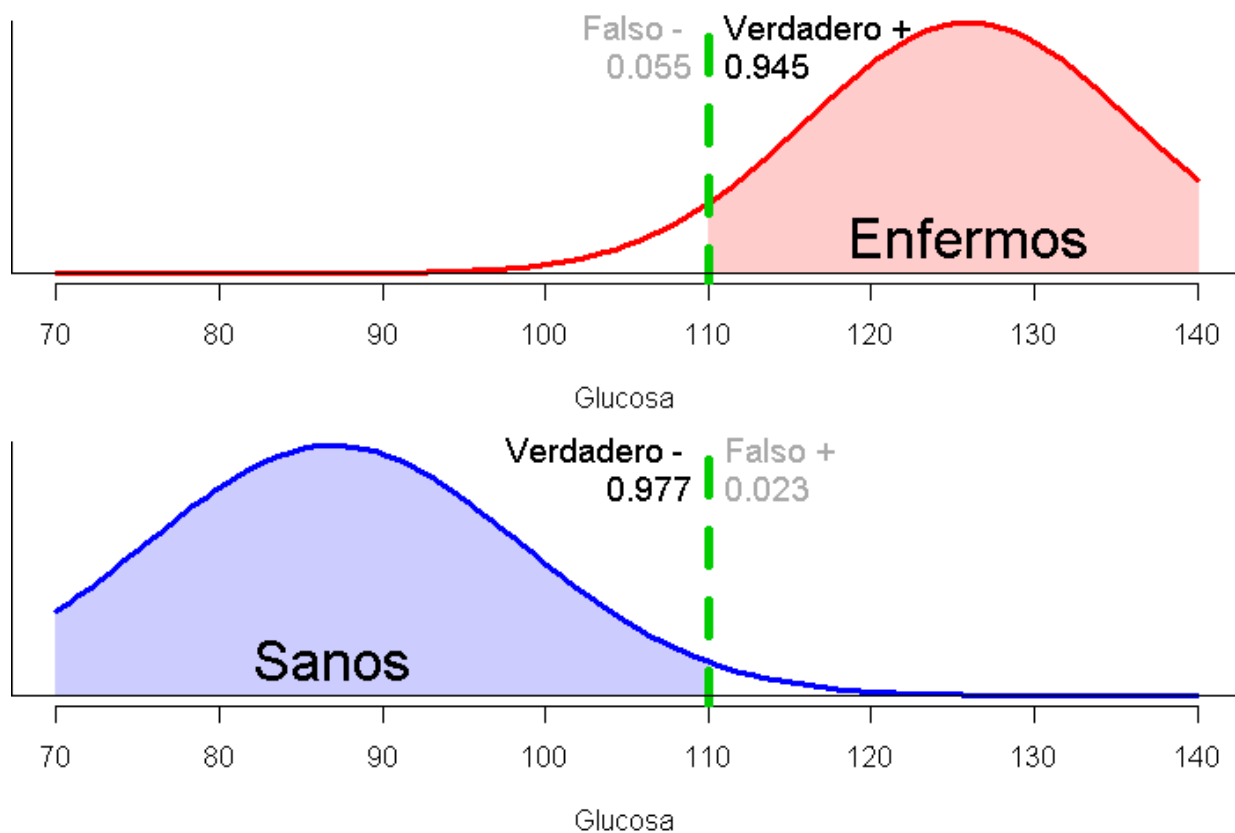
Prueba de Estudio	Prueba de oro Enfermos	Prueba de oro Sanos
+	400	50 positivos falsos
-	100 negativos falsos	450
	500	500

Prevalencia de 10%

Prueba de Estudio	Prueba de oro Enfermos	Prueba de oro Sanos
+	80	90 positivos falsos
-	20 negativos falsos	810
	100	900



Funcionamiento de la prueba diagnóstica de glucemia

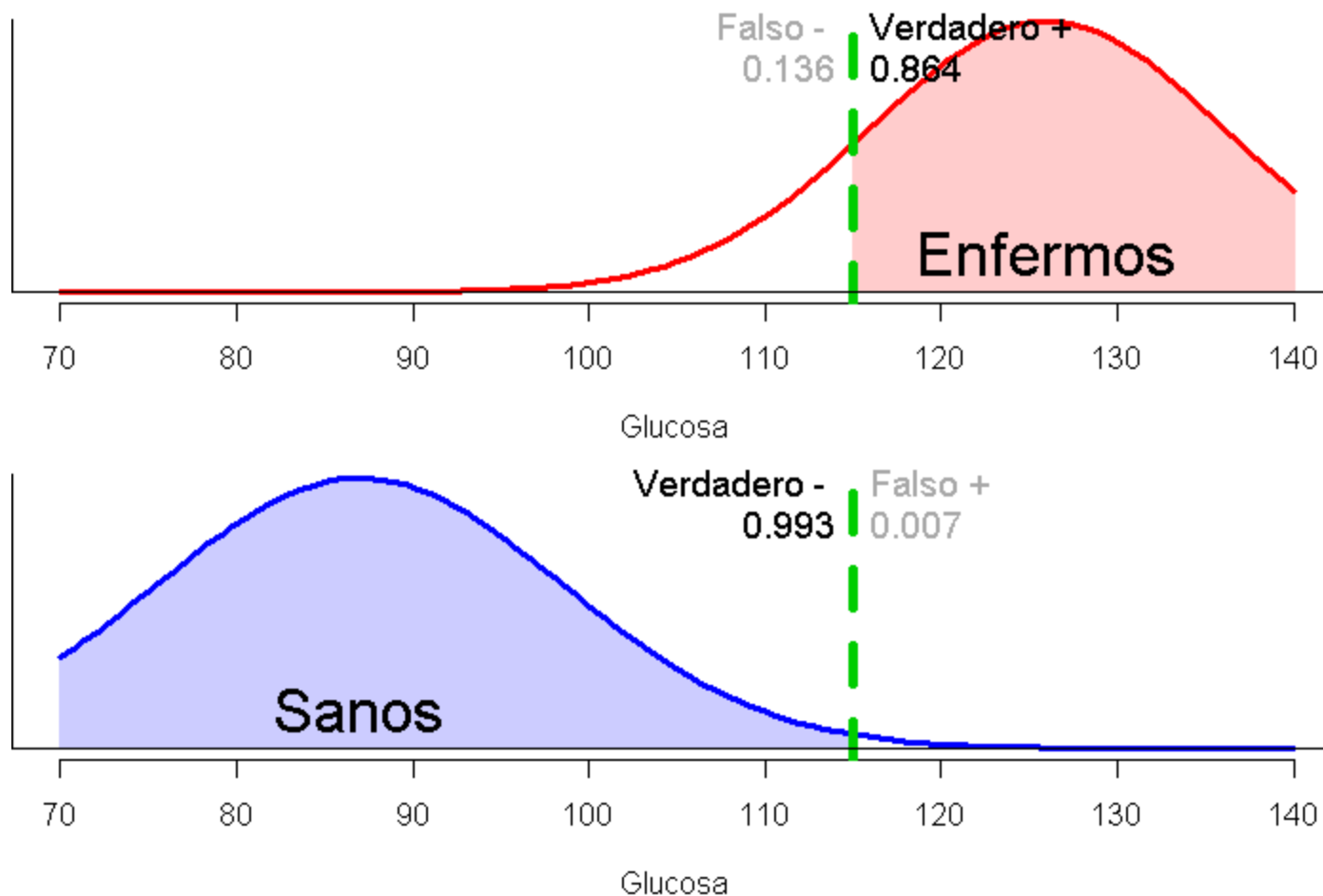


- **Valor límite: 110mg/dL**
 - Superior: test positivo.
 - Inferior: test negativo.

- **Probabilidad de acierto:**
 - Para enfermos
 - Verdadero positivo (sensibilidad)
 - Para sanos
 - Verdadero negativo (especificidad)

- **Probabilidad de error**
 - Para enfermos
 - Falso -
 - Para sanos
 - Falso +

¿Cómo definir el punto de corte de la prueba diagnóstica?



No es simple. **No es posible aumentar sensibilidad y especificidad al mismo tiempo.** Hay que elegir una solución de compromiso: Aceptable sensibilidad y especificidad.

Valor predictivo de las pruebas positivas y negativas

- ▶ Si la prueba es positiva, ¿Cuál es la probabilidad de que un individuo tenga la enfermedad?
- ▶ Si la prueba es negativa, ¿Cuál es la probabilidad de que la padezca?
- ▶ Valor predictivo positivo: Proporción de individuos con una prueba positiva que tienen la enfermedad.
- ▶ Valor predictivo negativo: Proporción de individuos con una prueba negativa que no tienen la enfermedad.

Valor predictivo de las pruebas positivas y negativas

	Prueba de Estudio	Prueba de oro Enfermos	Prueba de oro Sanos
Valor predictivo positivo =	+	a = Individuos enfermos y positivos	b = Individuos sanos y positivos
$a / a + b$		c = Individuos enfermos y negativos	d = Individuos sanos y negativos
Valor predictivo negativo =	-	$a + c = \text{Total de Individuos enfermos}$	$b + d = \text{Total de Individuos sanos}$
$d / c + d$			



Valor predictivo

Prevalencia V P Positivo V P Negativo

720	10
180	90

90%

98,6%

33,3%

400	50
100	450

50%

88,9%

81,8%

80	90
20	810

10%

47,1%

97,5%

8	99
889	2

1%

7,5%

99,7%



Ejemplo: Tamizar 1.000 personas
Especificidad de la prueba = 99,6% (4/1000)

Prevalencia de VIH = 10%

Verdaderos positivos: 100 **Falsos positivos: 4**

Valor predictivo positivo: $100/104 = 96\%$

Prevalencia de VIH = 0,4%

Verdadero positivo: 4 **Falso positivo: 4**

Valor predictivo positivo: $4/8 = 50\%$

Valor predictivo: análisis con una prueba

Especificidad de la prueba 99,6%

Prevalencia de VIH

Valor Predictivo Positivo

10%

96%

5%

91%

2%

80%

3%

86%

1%

67%

0,5%

50%

0,3%

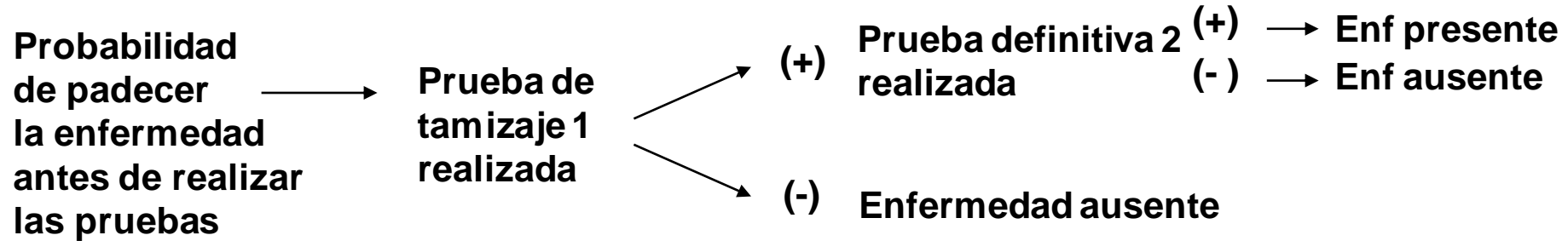
38%

0,1%

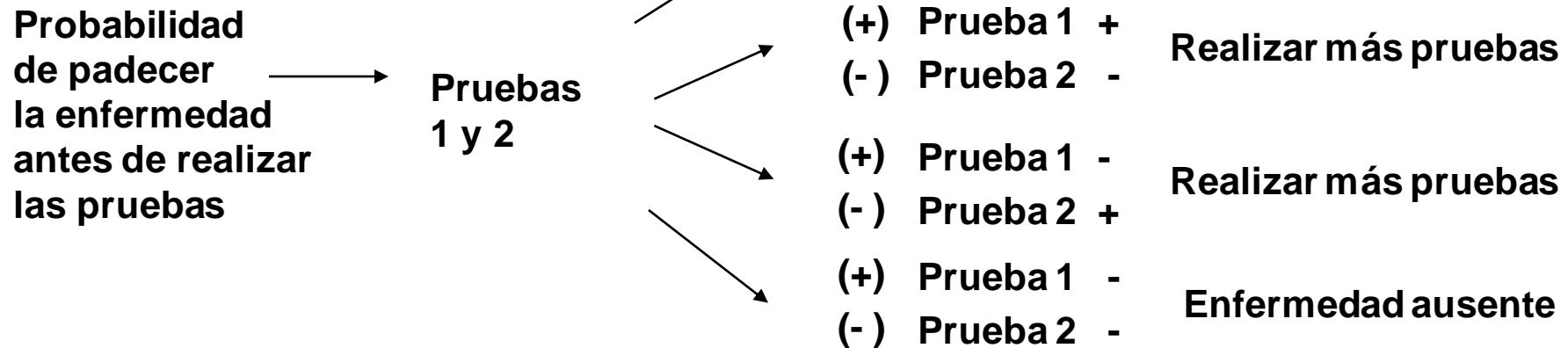
18%

Combinación de pruebas

Pruebas en serie



Pruebas en paralelo



Likelihood Ratio (Razón o Cociente de Verosimilitud)

- **LR +**
 - **Probabilidad de una prueba positiva en presencia de la enfermedad / Probabilidad de una prueba positiva en ausencia de enfermedad. Expresa cuantas veces es mas (o menos) probable un resultado positivo en enfermos que en no enfermos**
 - **= $S/(1 - E)$**
- **LR –**
 - **Probabilidad de una prueba negativa en presencia de la enfermedad / Probabilidad de una prueba negativa en ausencia de enfermedad**
 - **= $(1 - S) / E$**



Objetivo de usar la evidencia

- Promover la prestación de atención basada en la mejor evidencia disponible
- Incentivar a los proveedores de salud a que desarrollen habilidades que les permitan recolectar, evaluar e incorporar evidencia en la práctica clínica diaria
- Para poder lograr eso, los proveedores de salud deben comprender y evaluar intelectualmente los nuevos datos de la investigación clínica a medida que estén disponibles



En un mundo ideal

- Se sabría cuál es la atención más eficaz para cada condición
- Cada profesional de salud sabría cual es la atención más eficaz para cada paciente
- Cada profesional de salud prestaría la **atención más eficaz de la que tuviera conocimientos**

¿Cómo calcular sensibilidad, especificidad y valores predictivos?

Pruebas Diagnosticas

Seleccione el número de niveles de la prueba diagnóstica

2 Niveles | 3 Niveles | 4 Niveles

	Enfermedad		Totales
	Presente	Ausente	
Test Positivo	<input type="text" value="0"/>	<input type="text"/>	0
Test Negativo	<input type="text"/>	<input type="text"/>	0
	0	0	0

Analizar

Acerca de

Salir

Modulo de pruebas diagnósticas de la Unidad de Bioestadística del Hospital Ramón y Cajal, Madrid, España

Supongan

- En un grupo de 300 pacientes, 50 tienen meningitis por meningococo (*Neisseria meningitidis*) y 250 están libres de la enfermedad.
- Se realiza un cultivo mejorado en evaluación que logra detectar de los 50 que tienen meningitis, 40 como positivos en la prueba.
- El cultivo mejorado en evaluación que identificó como negativos a 230 de los sujetos que no tienen la enfermedad.

	Meningitis+	Sin meningitis	Total
Cultivo+	40	20	60
Cultivo-	10	230	240
Total	50	250	300

Supongan

- ¿Cuál es la sensibilidad del cultivo?
- ¿Su especificidad?
- El valor predictivo positivo
- El valor predictivo negativo

	Meningitis+	Sin meningitis	Total
Cultivo+	40	20	60
Cultivo-	10	230	240
Total	50	250	300



A calcular

	Meningitis+	Sin meningitis	Total
Cultivo+	40	20	60
Cultivo-	10	230	240
Total	50	250	300

Pruebas Diagnosticas

Seleccione el número de niveles de la prueba diagnóstica

2 Niveles | 3 Niveles | 4 Niveles

	Enfermedad		Totales
	Presente	Ausente	
Test Positivo	40	20	60
Test Negativo	10	230	240
	50	250	300

Analizar

Acerca de

Salir





Evaluación



Resultados

Indice	Valor	Intervalo Confianza 95%	
Sensibilidad	80,00 %	68,91 % a 91,09 %	
Especificidad	92,00 %	88,64 % a 95,36 %	
Valor predictivo positivo	66,67 %	54,74 % a 78,59 %	
Valor predictivo negativo	95,83 %	93,31 % a 98,36 %	
Cociente de probabilidad positivo	10,00	1	No aplicable
		2	6,42 a 15,57
Cociente de probabilidad negativo	0,22	1	No aplicable
		2	0,12 a 0,38

[1] Método de Miettinen [2] Método de Taylor



Concordancia

- ¿Qué tanto concuerda el cultivo con el diagnóstico de referencia de la enfermedad? (supongamos PCR para *N. meningitidis* en LCR)

	Meningitis+	Sin meningitis	Total
Cultivo+	40	20	60
Cultivo-	10	230	240
Total	50	250	300



Concordancia

- ¿Qué tanto concuerda el cultivo con el diagnóstico de referencia de la enfermedad? (supongamos PCR para *N. meningitidis* en LCR)

	Meningitis+	Sin meningitis	Total
Cultivo+	40	20	60
Cultivo-	10	230	240
Total	50	250	300

$40+230=270$
 $270/300*100=90\%$
 Concordancia absoluta

$10+20=30$
 $30/300*100=10\%$
 Disconcordancia absoluta

Coeficiente Kappa

- Compara los valores de dos variables nominales tales que sus valores pueden ser los mismos
- Tablas cuadradas
- Mide el grado de acuerdo entre las dos variables
- Asume valores entre 0 y 1
- Valores próximos a 1 : total acuerdo.
- Valores próximos a 0 : total desacuerdo

EVALUACIÓN DE LOS MÉTODOS DE GRAHAM Y PIN TAPE EN EL DIAGNÓSTICO DE *Enterobius vermicularis*

María Beltrán F¹, Hara T², Raúl Tello C¹

RESUMEN

El objetivo del estudio fue evaluar el método de *pin tape* (PT) respecto al método de Graham (MG) para el diagnóstico e intensidad de infección por *E. vermicularis* en niños. Se obtuvieron 250 muestras de 125 niños de ambos sexos, comprendidos entre 3 y 13 años de edad. Con el método de Graham se tuvo una prevalencia de 47% y con el *pin tape* 33,6%, asimismo, se detectó mayor intensidad de infección (número de huevos por lámina) con el MG que con el PT. La sensibilidad del PT respecto al MG fue de 71%, especificidad de 100% y concordancia (Kappa) 0,72. Se recomienda el uso del método de Graham respecto al *pin tape*, debido a que el método de Graham resulta más práctico, tiene mayor sensibilidad y detecta mayor número de huevos por lámina.

Palabras clave: *Enterobius vermicularis*; Diagnóstico; Niños (fuente: DeCS BIREME).

Tabla 1. Evaluación del total de casos según método de estudio. Método de Graham frente a *Pin Tape*.

		Método de Graham		Total
		Positivo	Negativo	
Pin Tape	Positivo	42	0	42
	Negativo	17	66	83
Total		59	66	125

DISCUSIÓN

Se encontró una buena concordancia diagnóstica entre los métodos de Graham y *Pin Tape* con un índice Kappa de 0,72; sin embargo, con el *Pin Tape* se detectaron menos casos y su uso resulta poco práctico debido a que debe permanecer durante toda la noche, por lo tanto hay que acudir dos veces al mismo lugar para obtener una muestra, a diferencia del método de

Tabla 1. Evaluación del total de casos según método de estudio. Método de Graham frente a *Pin Tape*.

		Método de Graham		Total
		Positivo	Negativo	
Pin Tape	Positivo	42	0	42
	Negativo	17	66	83
Total		59	66	125

Implicaciones Estadísticas

- un valor de concordancia diagnóstica, evaluado con la prueba Kappa, igual a 0,72 es alto
- sería importante conocer el valor de la p ,
- el cual puede variar de acuerdo con el tamaño de la muestra (n).



Tabla de contingencia VAR00001 * VAR00002

Recuento

		VAR00002		Total
		,00	1,00	
VAR00001	,00	3	1	4
	1,00	0	3	3
Total		3	4	7

Medidas simétricas

		Valor	Error típ. asint. ^a	T aproximada ^b	Sig. aproximada
Medida de acuerdo	Kappa	,720	,249	1,984	,047
N de casos válidos		7			

a. Asumiendo la hipótesis alternativa.

b. Empleando el error típico asintótico basado en la hipótesis nula.

Tabla 1. Evaluación del total de casos según método de estudio. Método de Graham frente a *Pin Tape*.

		Método de Graham		Total
		Positivo	Negativo	
Pin Tape	Positivo	42	0	42
	Negativo	17	66	83
Total				

Concordancia entre dos observadores con dos o más categorías

Origen de datos | Resultados

Número de categorías: 2 | Nivel de confianza (%): 95,0

Entrada de datos

Clasificaciones

		Observador 2	
		1	2
Observador 1	1	42	0
	2	17	66

Kappa	EE	IC (95,0%)	
0,7229	0,0600	0,6053	0,8406

Kappa mínimo: -0,0730

Kappa máximo: 0,7329

Prueba de significación

Estadístico Z	Valor p
8,4118	0,0000

(p<0,0001)



Tabla 1. Evaluación del total de casos según método de estudio. Método de Graham frente a *Pin Tape*.

		Método de Graham		Total
		Positivo	Negativo	
Pin Tape	Positivo	42	0	42
	Negativo	17	66	83
Total				

Pruebas diagnósticas simples

Origen de datos y resultados

Opciones

Datos tabulados

Valores predictivos

Nivel de confianza (%)

Prueba de referencia

	Positivo	Negativo	Total
Prueba diagnóstica	42	0	42
	17	66	83
Total	59	66	125



Pruebas diagnósticas simples

Origen de datos y resultados

Opciones

Datos tabulados

Valores predictivos

Nivel de confianza (%)

		Prueba de referencia		
		Positivo	Negativo	Total
Prueba diagnóstica	Positivo	42	0	42
	Negativo	17	66	83
Total		59	66	125

	Valor	IC (95%)	
Sensibilidad (%)	71,19	58,78	83,59
Especificidad (%)	100,00	99,24	100,00
Índice de validez (%)	86,40	79,99	92,81
Valor predictivo + (%)	100,00	98,81	100,00
Valor predictivo - (%)	79,52	70,23	88,80
Prevalencia (%)	47,20	38,05	56,35



- El incluir la significancia estadística de los resultados encontrados y las diferencias, especialmente cuando se están comparando pruebas diagnósticas, permite ponderar mejor los hallazgos encontrados y tomar mejores decisiones sobre su utilidad, por lo cual esto siempre debe ser considerado en la presentación de resultados.



BIOMETRICS 33, 159–174
March 1977

The Measurement of Observer Agreement for Categorical Data

J. RICHARD LANDIS

Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109 U.S.A.

GARY G. KOCH

Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina 27514 U.S.A.



AGREEMENT MEASURES FOR CATEGORICAL DATA

165

Kappa Statistic

<0.00
0.00–0.20
0.21–0.40
0.41–0.60
0.61–0.80
0.81–1.00

Strength of Agreement

Poor
Slight
Fair
Moderate
Substantial
Almost Perfect

Although these divisions are clearly arbitrary, they do provide useful “benchmarks” for the discussion of the specific example in Table 1.



