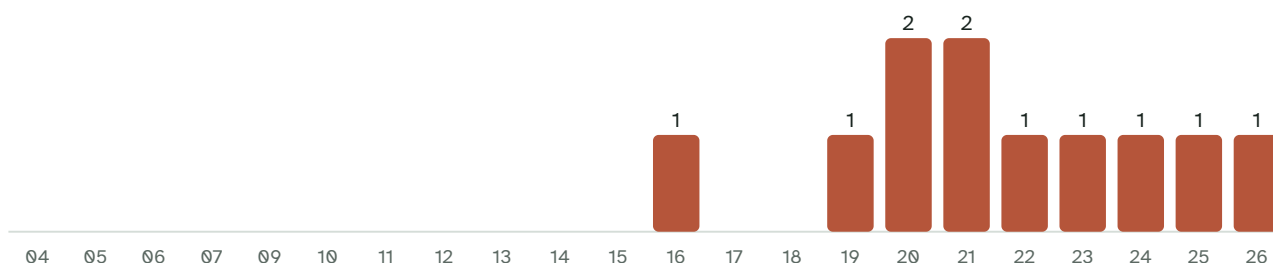


Validez de pruebas diagnósticas

Validar una prueba diagnóstica es compararla con un patrón de referencia y resumir sus aciertos y errores en una tabla 2x2. De esa tabla salen la sensibilidad y la especificidad (validez), los valores predictivos (seguridad), los cocientes de probabilidad y, cuando el resultado es continuo, la curva ROC. En el EIR cae todos los años desde 2019: fórmulas de la tabla, qué parámetro depende de la prevalencia, qué pasa al mover el punto de corte y qué mide el área bajo la curva.

Apuntes para el EIR · versión 2026-09-27 · **sin revisión enfermera**. Todas las preguntas son reales, de exámenes EIR, con su año y su número; ninguna está escrita por IA. Las respuestas están en el solucionario, al final.

Cómo cae en el EIR



11 preguntas de este tema en 22 exámenes (2004-2026, ordinarias, clasificación segura).

Lo más repetido

- Valor predictivo positivo: definición e interpretación (2 exámenes)
- Curva ROC y área bajo la curva: capacidad discriminativa (2 exámenes)
- Valores predictivos: dependen de la prevalencia (alta prevalencia aumenta VPP) (2 exámenes)

Índice

Resumen para repasar	3
Puntos clave	5
1. Validar una prueba: patrón de referencia y tabla 2x2	6
2. Sensibilidad y especificidad: la validez de la prueba	9
3. El punto de corte: sensibilidad frente a especificidad	12
4. Valores predictivos: la seguridad de la prueba	14
5. La prevalencia cambia los valores predictivos	16
6. Cocientes de probabilidad e índice de Youden	18
7. Curvas ROC y área bajo la curva	19
8. Pruebas en serie y en paralelo, y calidad de los estudios	21
Todas las preguntas del tema (11)	22
Solucionario	25
Fuentes	26

Resumen para repasar

1. Validar una prueba: patrón de referencia y tabla 2x2

- Un test debe tener **validez** (mide lo que debe medir: sensibilidad y especificidad), **reproductividad** (mismos resultados al repetirlo) y **seguridad** (valores predictivos) ^{F1}.
- Se valida frente al estado real o la **prueba de referencia (gold standard)** ^{F1}; rara vez existe un método de exactitud absoluta ^{F3}.
- Tabla 2x2: prueba en filas, enfermedad en columnas; **a** = VP, **b** = FP, **c** = FN, **d** = VN ^{F1,F3}.
- Sensibilidad **$a/(a+c)$** , especificidad **$d/(b+d)$** , VPP **$a/(a+b)$** , VPN **$d/(c+d)$** ^{F3}; **$c/(a+c)$** es la fracción de falsos negativos ^{F1,F3}.
- **Valor global = $(VP + VN)/total$** : aciertos entre todos; mismo peso a sensibilidad y especificidad, valor muy limitado ^{F4}.

2. Sensibilidad y especificidad: la validez de la prueba

- **Sensibilidad**: probabilidad de clasificar bien a un enfermo (fracción de verdaderos positivos) ^{F1,F2}.
- **Especificidad**: probabilidad de clasificar bien a un sano (fracción de verdaderos negativos = $1 - FFP$) ^{F1,F2}.
- Son **propiedades intrínsecas**, independientes de la prevalencia ^{F1,F4}.
- **Alta sensibilidad** para el **cribado** y enfermedades peligrosas pero tratables ^{F1,F4}.
- **Alta especificidad** para **confirmar** antes de tratar y cuando un falso positivo acarrea graves consecuencias físicas, psicológicas o económicas ^{F1,F4}.

3. El punto de corte: sensibilidad frente a especificidad

- En pruebas continuas se elige un **valor de corte**; sanos y enfermos se solapan y el límite normal-anormal es en general **arbitrario** ^{F2,F3}.
- **Criterios menos estrictos** (bajar el umbral si lo patológico son valores altos) → ↑ **sensibilidad**, ↓ **especificidad**, más falsos positivos; más estrictos → lo contrario ^{F3}.
- Diabetes: la glucemia basal diagnóstica se rebajó a **$\geq 7,0$ mmol/l (126 mg/dl)** ^{F7,F8}; en ARIC, un umbral más bajo de glucemia basal alterada (para predecir diabetes futura) dio 70 % de sensibilidad frente a 50 %, a costa del doble de anormales ^{F8}.
- Sensibilidad, especificidad y valores predictivos son **probabilidades**; el **límite patológico no lo es** ^{F3}.
- La **fiabilidad** (resultados coherentes al repetir) es una propiedad distinta de la validez ^{F1,F3}.

4. Valores predictivos: la seguridad de la prueba

- Responden a la pregunta del clínico: con este resultado, ¿está enfermo? ^{F1,F4}.
- **VPP**: proporción de **positivos que están realmente enfermos** (VP entre todos los positivos) ^{F1,F3}.
- **VPN**: proporción de negativos que están realmente sanos ^{F1,F3}.
- Tacto rectal: VPP 70,21 %, VPN 71,98 % ^{F1}.
- VPP del 90 % = el 90 % de los positivos están enfermos; no es la sensibilidad ni el valor global ^{F1,F4}.

5. La prevalencia cambia los valores predictivos

- Los valores predictivos dependen sobre todo de la **prevalencia** ^{F3}.
- **Prevalencia alta** → ↑ **VPP**; prevalencia baja → VPP bajo y VPN alto ^{F1,F4}.
- VIH con sensibilidad y especificidad del 99,5 %: VPP **29,9 %** con prevalencia del 0,21 % y **98,7 %** con el 28,6 % ^{F1}.
- COVID-19: con prevalencia del 1 %, incluso con 99 % de especificidad **la mitad de los positivos serían falsos** ^{F5}; si la prevalencia cambia entre olas, cambian los **valores predictivos** ^{F5}, no la sensibilidad ni la especificidad ^{F1}.

6. Cocientes de probabilidad e índice de Youden

- **CP+** = sensibilidad/(1 – especificidad); **CP–** = (1 – sensibilidad)/especificidad ^{F1}.
- **No dependen de la prevalencia**: sirven para comparar pruebas ^{F1,F4}.
- **Índice de Youden** = sensibilidad + especificidad – 1 ^{F4}.

7. Curvas ROC y área bajo la curva

- Curva ROC: pares (**1 – especificidad, sensibilidad**) para todos los cortes ^{F1,F2}.
- **Eje Y: sensibilidad; eje X: falsos positivos (1 – especificidad)** ^{F2}.
- **Diagonal = no discrimina**; discriminación perfecta = pasa por el **vértice superior izquierdo** ^{F2}.
- **Área bajo la curva**: capacidad discriminativa global; **0,5 = azar, 1 = perfecta**; independiente de la prevalencia ^{F1,F2,F4}.
- El **kappa** mide **concordancia** entre observadores, no discriminación ^{F9}.

8. Pruebas en serie y en paralelo, y calidad de los estudios

- **En paralelo**: diagnostica más enfermos pero más sanos como enfermos; **en serie**: pocos falsos positivos pero se escapan enfermos ^{F1}.
- Los índices son estimaciones: con **intervalo de confianza** y cuidando muestreo y criterio de referencia para evitar sesgos ^{F1,F2}.

Puntos clave

- 1 Sensibilidad = $a/(a+c)$: probabilidad de que la prueba sea positiva en los enfermos; especificidad = $d/(b+d)$: probabilidad de que sea negativa en los sanos ^{F3}.
- 2 VPP = $a/(a+b)$: verdaderos positivos entre todos los positivos; VPN = $d/(c+d)$ ^{F3}.
- 3 Valor global = $(VP + VN) / \text{total}$: proporción de sujetos bien clasificados; da el mismo peso a sensibilidad y especificidad y tiene valor limitado ^{F4}.
- 4 Cribado: alta sensibilidad. Confirmación antes de tratar, o cuando un falso positivo acarrea graves consecuencias: alta especificidad ^{F1,F4}.
- 5 Criterios menos estrictos para considerar positiva la prueba (por ejemplo, bajar el umbral cuando lo patológico son los valores altos) aumentan la sensibilidad y reducen la especificidad ^{F3}.
- 6 Los valores predictivos dependen de la prevalencia: si sube, aumenta el VPP; la sensibilidad, la especificidad y los cocientes de probabilidad no cambian ^{F1,F3}.
- 7 Curva ROC: sensibilidad en el eje Y y $1 - \text{especificidad}$ (falsos positivos) en el eje X; la diagonal es una prueba que no discrimina y la perfecta pasa por el vértice superior izquierdo ^{F2}.
- 8 El área bajo la curva ROC mide la capacidad discriminativa: 0,5 = azar, 1 = perfecta ^{F2,F4}.

1. Validar una prueba: patrón de referencia y tabla 2x2

Una buena prueba diagnóstica es la que da resultados positivos en los enfermos y negativos en los sanos. Pita Fernández y Pértegas Díaz resumen en tres las condiciones que deben exigirse a un test ^{F1}:

- **Validez:** el grado en que un test mide lo que se supone que debe medir; **la sensibilidad y la especificidad son medidas de su validez** ^{F1}.
- **Reproductividad:** la capacidad del test para ofrecer los mismos resultados cuando se repite su aplicación en circunstancias similares; la determinan la variabilidad biológica, la del observador y la del propio test ^{F1}.
- **Seguridad:** viene determinada por el **valor predictivo** de un resultado positivo o negativo, y esta probabilidad está muy influida por la prevalencia ^{F1}.

Además conviene que el test sea sencillo de aplicar, aceptado por los pacientes, con los mínimos efectos adversos y económicamente soportable ^{F1}. Para una prueba de detección sistemática (cribado), la OMS pide que sea de bajo costo, fácil de aplicar, aceptable, **fiable** (resultados coherentes) y **válida** (clasifica correctamente a las personas en un grupo con enfermedad y otro sin ella) ^{F3}.

Para validar una prueba se compara su resultado con el **estado real** del paciente o, en su defecto, con el resultado de la **prueba de referencia o «gold standard»** (patrón oro) ^{F1}. Muchas veces el diagnóstico real de todos los individuos es desconocido, y se toma como comparador la prueba de referencia usada hasta la aparición de la nueva prueba ^{F4}. Rara vez se dispone de un método de exactitud absoluta, y las pruebas muy exactas a menudo son caras y cruentas; por eso en la práctica se usan pruebas más simples, cuya validez hay que conocer ^{F3}.

En el caso más sencillo, una **prueba dicotómica** (positivo/negativo), los sujetos estudiados se reparten en cuatro grupos en una **tabla 2x2** que enfrenta el resultado de la prueba (**en filas**) con el verdadero diagnóstico (**en columnas**) ^{F1}. El resultado puede ser correcto (verdadero positivo y verdadero negativo) o incorrecto (falso positivo y falso negativo) ^{F1}. La OMS rotula las casillas con letras ^{F3}:

Resultado de la prueba	Enfermedad presente	Enfermedad ausente	Total
Positiva	a = verdaderos positivos (VP)	b = falsos positivos (FP)	a + b
Negativa	c = falsos negativos (FN)	d = verdaderos negativos (VN)	c + d
Total	a + c (enfermos)	b + d (sanos)	a + b + c + d F3

Con esa tabla se calculan todos los índices. Los cuatro primeros los da la OMS con sus letras (cuadro 6.5), y cada uno se define como una **probabilidad** ^{F3}; el valor global lo da Cirugía Española como aciertos entre todos ^{F4}.

Índice	Fórmula	Qué probabilidad expresa
Sensibilidad	$a / (a + c) = VP / (VP + FN)$	Que la prueba sea positiva en las personas que tienen la enfermedad ^{F3}
Especificidad	$d / (b + d) = VN / (VN + FP)$	Que la prueba sea negativa en las personas que no tienen la enfermedad ^{F3}
Valor predictivo positivo	$a / (a + b) = VP / (VP + FP)$	Que la persona tenga la enfermedad cuando la prueba da positivo ^{F3}
Valor predictivo negativo	$d / (c + d) = VN / (VN + FN)$	Que la persona no tenga la enfermedad cuando la prueba da negativo ^{F3}
Valor (predictivo) global	$(VP + VN) / \text{total} = (a + d) / (a + b + c + d)$	Aciertos entre todos los sujetos estudiados ^{F4}



Cómo leer la tabla sin equivocarse

Sensibilidad y especificidad se calculan **por columnas** (sobre enfermos, $a + c$, y sobre sanos, $b + d$); los valores predictivos, **por filas** (sobre positivos, $a + b$, y sobre negativos, $c + d$) ^{F3}. Si en el denominador está $a + c$, el numerador que corresponde a la sensibilidad es **a**; $c / (a + c)$ es la proporción de enfermos con resultado negativo, la fracción de falsos negativos ($1 - \text{sensibilidad}$) ^{F1,F3}.

El **valor predictivo global** (en inglés *overall accuracy*; en muchos textos, simplemente «valor global») resume en una sola cifra la validez de la prueba: es la proporción de **aciertos** ($VP + VN$) sobre **todos** los sujetos estudiados ^{F4}. En el ejemplo del tacto rectal de la sección 2 serían $(634 + 1.251) / 2.641 \approx 71,4 \%$ (cálculo con los datos de la tabla 2 de Fisterra) ^{F1,F4}.

El valor global tiene un valor muy limitado

Da **el mismo peso a la sensibilidad y a la especificidad**. Sancho-Insenser y González-Castillo lo consideran, pese a su nombre, el parámetro menos objetivo para describir la validez de una prueba dicotómica; es claramente inferior a las razones de verosimilitud y, según ellos, debería estar en desuso ^{F4}. No confundirlo con la prevalencia (enfermos entre el total, $(a + c) / \text{total}$) ni con la proporción de positivos $((a + b) / \text{total})$, que no miden aciertos (se leen directamente en la tabla 2×2) ^{F3}.

Se ha realizado un estudio de validación de una prueba diagnóstica, cuyos resultados se reflejan en una...: La sensibilidad se calcula mediante la fórmula: $a / (a+c)$

Se ha realizado un estudio de validación de una prueba diagnóstica, cuyos resultados se reflejan en una tabla de contingencia. ¿Cuál de las siguientes afirmaciones es la correcta?:

1. El valor global se calcula mediante la fórmula: $a / (a+b+c+d)$.
2. La especificidad se calcula mediante la fórmula: $c / (a+c)$.
3. La sensibilidad se calcula mediante la fórmula: $a / (a+c)$.
4. El Valor Predictivo Negativo se calcula mediante la fórmula: $(b+c) / (a+b+c+d)$.

Cuando se habla de valor global en el proceso de toma de decisiones diagnósticas, hace referencia a: La proporción de sujetos correctamente clasificados por el test en relación al total...

Cuando se habla de valor global en el proceso de toma de decisiones diagnósticas, hace referencia a:

1. El número de enfermos que existe en una población determinada en un período de tiempo concreto.
2. Es el número total de positivos obtenidos en el proceso diagnóstico en relación al total de sujetos que componen la muestra estudiada.
3. La proporción de sujetos correctamente clasificados por el test en relación al total de sujetos que componen la muestra de estudio.
4. Grado en que los resultados de un estudio son generalizables a otros sujetos o poblaciones.

2. Sensibilidad y especificidad: la validez de la prueba

Sensibilidad: es la **probabilidad de clasificar correctamente a un individuo enfermo**, es decir, la probabilidad de que un sujeto enfermo obtenga un resultado positivo; es la **capacidad del test para detectar la enfermedad** ^{F1}. Se estima como la proporción de enfermos con resultado positivo, y por eso se llama también **fracción de verdaderos positivos (FVP)** ^{F1,F2}.

Especificidad: es la **probabilidad de clasificar correctamente a un individuo sano**, es decir, la probabilidad de que un sujeto sano obtenga un resultado negativo; es la **capacidad para detectar a los sanos** ^{F1}. Se llama también **fracción de verdaderos negativos (FVN)** y es igual a **1 – la fracción de falsos positivos (FFP)** ^{F1,F2}.

Ejemplo de Fistera: 2.641 pacientes con sospecha de cáncer de próstata; se comparó el tacto rectal (normal o anormal) con la biopsia (prueba de referencia) ^{F1}. Hubo 1.121 cánceres (42,45 %). La **sensibilidad** del tacto fue del **56,56 %** (634/1.121) y la **especificidad** del **82,3 %** (1.251/1.520) ^{F1}. Por tanto, un 43,44 % de los pacientes que tenían cáncer presentaban un tacto normal (falsos negativos) ^{F1}.

Tacto rectal	Cáncer (biopsia)	Patología benigna	Total
Anormal	634	269	903
Normal	487	1.251	1.738
Total	1.121	1.520	2.641 ^{F1}

Sensibilidad y especificidad son **propiedades intrínsecas de la prueba**: definen su validez **independientemente de la prevalencia** de la enfermedad en la población a la que se aplica ^{F1,F4}. Su inconveniente es que no responden a la pregunta del clínico ante un resultado concreto; esa pregunta la responden los valores predictivos (sección 4) ^{F1,F4}.

Lo ideal sería una prueba de alta sensibilidad y alta especificidad, pero no siempre es posible, y hay que elegir según el uso ^{F1}:

	Prueba muy SENSIBLE	Prueba muy ESPECÍFICA
Evita sobre todo	Los falsos negativos (no se escapan enfermos) ^{F1,F4}	Los falsos positivos (no se etiqueta de enfermo a un sano) ^{F1,F4}
Uso típico	Cribado: captar a todos los enfermos, aun a costa de algunos falsos positivos ^{F1,F4}	Confirmación del diagnóstico: antes de iniciar un tratamiento se suele pedir una prueba confirmatoria, idealmente de alta especificidad ^{F4}

	Prueba muy SENSIBLE	Prueba muy ESPECÍFICA
Cuándo es especialmente adecuada	Cuando no diagnosticar puede ser fatal: enfermedades peligrosas pero tratables (linfomas, tuberculosis); o cuando un falso positivo no produce serios trastornos psicológicos o económicos (mamografía en el cáncer de mama) ^{F1}	Enfermedades graves pero sin tratamiento que las haga curables ; cuando interesa mucho conocer la ausencia de enfermedad; o cuando diagnosticar a alguien de un mal que no padece puede acarrear graves consecuencias físicas, psicológicas o económicas (Fisterra pone el ejemplo del sida) ^{F1,F4}

La OMS lo plantea igual para el cribado: los criterios de una prueba dependen de las consecuencias de clasificar a individuos como falsos negativos y falsos positivos; en una enfermedad grave de recién nacidos puede preferirse una alta sensibilidad, aceptando el costo de muchos falsos positivos, que luego hay que seguir para distinguir quiénes son positivos verdaderos ^{F3}.



Tratamiento agresivo o invasivo: especificidad

Si el resultado positivo va a desencadenar un manejo **invasivo o con efectos secundarios graves**, un falso positivo expondría a un sano a ese daño. En esa situación se busca una prueba de **alta especificidad**: las pruebas confirmatorias deben serlo **para evitar falsos positivos**, sobre todo cuando un diagnóstico erróneo puede acarrear graves consecuencias físicas, psicológicas o económicas ^{F1,F4}.

Sensible para cribar, específica para confirmar

Sensibilidad → **Screening** (cribado) y enfermos que no deben escaparse. **Especificidad** → **Evitar falsos positivos** antes de tratar.

Dos «sensibilidades» distintas

Aquí la sensibilidad es la capacidad del test para **detectar la enfermedad** (fracción de verdaderos positivos) ^{F1}. No es la «sensibilidad al cambio» de un cuestionario o escala, que es otra propiedad psicométrica (se estudia en M26-T02 con la fiabilidad y la validez de los instrumentos).

Si sospechamos la presencia de una enfermedad severa, cuyo manejo conlleva medidas invasivas con...: Tener una elevada especificidad

Si sospechamos la presencia de una enfermedad severa, cuyo manejo conlleva medidas invasivas con potenciales efectos secundarios graves, ¿cuál sería la principal característica que buscaremos en la prueba diagnóstica a realizar?:

1. Tener una elevada sensibilidad.
2. Tener una elevada especificidad.
3. Tener un elevado valor predictivo positivo.
4. Tener un elevado valor predictivo negativo.

3. El punto de corte: sensibilidad frente a especificidad

Muchas pruebas no dan un resultado dicotómico sino un **valor numérico** (por ejemplo, una glucemia). Para clasificar a los pacientes hay que elegir un **valor de corte** y considerar positivos los valores situados a un lado de él ^{F1,F2}. Con cada punto de corte se obtiene un par distinto de sensibilidad y especificidad ^{F1,F2}.

Las distribuciones de los resultados en sanos y enfermos suelen **solaparse**: es imposible elegir un valor que separe netamente la normalidad de la anormalidad, y siempre hay personas sanas en el lado «anormal» y casos verdaderos en el lado «normal» ^{F3}. Por eso el equilibrio entre sensibilidad y especificidad se busca fijando **un límite entre lo normal y lo anormal que en general es arbitrario** ^{F3}.



Mover el punto de corte

Si se aplican **criterios menos estrictos** para considerar positiva una prueba, **aumenta la sensibilidad y disminuye la especificidad**; con **criterios más estrictos**, aumenta la especificidad pero disminuye la sensibilidad ^{F3}. Si se intenta aumentar la sensibilidad para incluir a todos los positivos verdaderos, necesariamente aumenta el número de **falsos positivos** ^{F3}. Siempre hay que buscar un equilibrio: **el incremento de una supone la reducción de la otra** ^{F3}.

En la curva ROC se ve el mismo compromiso: si se modifica el valor de corte para obtener mayor sensibilidad, solo puede hacerse a expensas de disminuir la especificidad ^{F2}. Elegir el valor de corte exige conocer los riesgos y beneficios de las decisiones que se derivan del resultado (el coste de un falso positivo frente al de un falso negativo) y la prevalencia ^{F2}.

Ejemplo de la diabetes. Aplicando esa regla general ^{F3}: cuando lo patológico son los valores altos, como en la glucemia, **bajar la cifra umbral** es un criterio **menos estricto** para considerar positiva la prueba; por tanto aumenta la sensibilidad (se detectan más enfermos) y disminuye la especificidad (más falsos positivos). La consulta de la OMS de 1998 (informe provisional, en paralelo con el comité de expertos de la Asociación Americana de Diabetes) **propuso rebajar el valor diagnóstico de la glucemia plasmática en ayunas a $\geq 7,0$ mmol/l** ^{F7}; es el criterio que la OMS y la FID mantuvieron en 2006: glucemia en ayunas $\geq 7,0$ mmol/l (126 mg/dl) o a las 2 horas $\geq 11,1$ mmol/l (200 mg/dl) ^{F8}.

El mismo informe OMS/FID muestra el precio de ese aumento de sensibilidad con otro umbral, el de la glucemia basal alterada usado para **predecir diabetes futura** (no para diagnosticarla): en el estudio ARIC, una glucemia en ayunas $\geq 5,6$ mmol/l tuvo una sensibilidad del 70 % para predecir diabetes, frente al 50 % con $\geq 5,9$ mmol/l, pero a costa de clasificar como anormal al 40 % de la población en lugar del 20 % ^{F8}.

La Asociación Americana de Diabetes en 1997 y la Organización Mundial de la Salud en 1998 aprobaron nuevos...:
Incrementar la sensibilidad de la prueba

La Asociación Americana de Diabetes en 1997 y la Organización Mundial de la Salud en 1998 aprobaron nuevos criterios para el diagnóstico de la Diabetes Mellitus. Se recomendó el descenso del valor de glucemia basal de 140 mg/dl a 126 mg/dl, para diagnosticar a una persona de Diabetes. Este cambio de criterio consigue:

1. Incrementar la sensibilidad de la prueba.
2. Incrementar la especificidad de la prueba.
3. Disminuir los falsos positivos de la prueba.
4. Incrementar la fiabilidad de la prueba.

Características probabilísticas y punto de corte. La sensibilidad, la especificidad y los valores predictivos positivo y negativo se definen como **probabilidades** (de un resultado según el estado real, o de un estado real según el resultado) ^{F1,F3}. El límite entre lo normal y lo patológico, en cambio, **no es una probabilidad**: es un valor de la escala de la prueba que se fija, en general de forma arbitraria, para dicotomizar los resultados ^{F3}.

Validez no es fiabilidad

La **fiabilidad** o reproductividad es que el test dé los mismos resultados al repetirlo en circunstancias similares ^{F1,F3}; la **validez**, que clasifique correctamente a sanos y enfermos, y se mide con la sensibilidad y la especificidad ^{F1,F3}. Lo que mueve el punto de corte, según la OMS, es el balance entre sensibilidad y especificidad ^{F3}: es decir, la validez. Ninguna de las fuentes vincula el punto de corte con la repetibilidad del test.

Al realizar una prueba diagnóstica, ¿cuáles de las siguientes opciones no es una característica probabilística: Límites patológicos

Al realizar una prueba diagnóstica, ¿cuáles de las siguientes opciones no es una característica probabilística?:

1. Sensibilidad.
2. Valor predictivo negativo.
3. Especificidad.
4. Límites patológicos.

4. Valores predictivos: la seguridad de la prueba

Sensibilidad y especificidad informan de la probabilidad de un resultado **en función de la verdadera condición** del paciente. Pero cuando se hace una prueba no se conoce el diagnóstico real, y la pregunta del clínico es la contraria: **ante un resultado positivo (negativo), ¿qué probabilidad hay de que el paciente esté realmente enfermo (sano)?** La responden los valores predictivos ^{F1,F4}.

- **Valor predictivo positivo (VPP):** la probabilidad de padecer la enfermedad si se obtiene un resultado positivo; se estima como la **proporción de pacientes con resultado positivo que finalmente resultaron estar enfermos**, es decir, **verdaderos positivos entre todos los positivos**, $a / (a + b)$ ^{F1,F3}.
- **Valor predictivo negativo (VPN):** la probabilidad de que un sujeto con resultado negativo esté realmente sano; se estima dividiendo los **verdaderos negativos entre el total de resultados negativos**, $d / (c + d)$ ^{F1,F3}.

La OMS lo define igual en su guía de pruebas de antígenos del SARS-CoV-2: el VPP es la probabilidad de que los pacientes con un resultado positivo tengan la enfermedad ^{F5}.

En el ejemplo del tacto rectal, el **VPP fue del 70,21 %** (634/903) y el **VPN del 71,98 %** (1.251/1.738): en el 70,21 % de los pacientes con tacto anormal se confirmó el cáncer, y de los que tuvieron un tacto normal, el 71,98 % estaban efectivamente sanos ^{F1}.

Si dicen «el 90 %...»	Parámetro	Denominador
...de los enfermos dan positivo	Sensibilidad	Enfermos (a + c) ^{F1,F3}
...de los sanos dan negativo	Especificidad	Sanos (b + d) ^{F1,F3}
...de los que dan positivo están realmente enfermos	Valor predictivo positivo	Positivos (a + b) ^{F1,F3}
...de los que dan negativo están realmente sanos	Valor predictivo negativo	Negativos (c + d) ^{F1,F3}
...de todos los resultados son aciertos, positivos y negativos	Valor global	Total (a + b + c + d) ^{F4}



Interpretar un VPP

Un VPP del 90 % significa que **el 90 % de los pacientes con un resultado positivo tienen realmente la enfermedad** ^{F1,F3}. No significa que la prueba detecte al 90 % de los enfermos (eso es la sensibilidad), ni que dé negativo en el 90 % de los sanos (la especificidad), ni que acierte el 90 % de todas las veces (el valor global) ^{F1,F4}.

Los valores predictivos tienen utilidad clínica y para informar al paciente, pero **dependen de la prevalencia** (sección 5); por eso Sancho-Insenser y González-Castillo los califican de ideales para el clínico, pero muy sensibles a valores extremos de la prevalencia y sin utilidad para comparar pruebas en poblaciones distintas ^{F1,F4}. Cuando se habla de «valor predictivo» para la proporción de verdaderos positivos entre todos los identificados como positivos, se trata del valor predictivo **positivo** ^{F1,F3}.

Valor predictivo positivo: definición e interpretación

En una prueba de cribado, la proporción de verdaderos positivos entre los identificados como positivos, se denomina:

1. Prevalencia.
2. Especificidad.
3. Sensibilidad.
4. Valor predictivo.

Valor predictivo positivo: definición e interpretación

En un estudio realizado en atención primaria para validar un test rápido de detección de antígenos del estreptococo β -hemolítico del grupo A en pacientes con faringitis, el valor predictivo positivo del test fue del 90 %. ¿Qué interpretación es correcta?:

1. Que el test identifica correctamente al 90 % de los pacientes con infección estreptocócica.
2. Que el 90 % de los pacientes con un resultado positivo en el test realmente presentan faringitis estreptocócica.
3. Que el test ofrece un resultado negativo en el 90 % de los pacientes sin infección.
4. Que el test acierta el 90 % de las veces, tanto en positivos como en negativos.

5. La prevalencia cambia los valores predictivos

El valor predictivo depende de la sensibilidad y la especificidad de la prueba y, **lo que es más importante, de la prevalencia** de la enfermedad en la población estudiada ^{F3}. Incluso con sensibilidad y especificidad elevadas, cuando la prevalencia es baja el VPP puede ser muy bajo ^{F3}. La prevalencia de los pacientes a los que se aplica la prueba puede ser muy distinta de la del estudio publicado que estableció su utilidad ^{F3}.

Prevalencia	VPP	VPN
Baja	Bajo: un positivo no permite confirmar el diagnóstico ^{F1}	Alto: un negativo descarta la enfermedad con mayor seguridad ^{F1,F4}
Alta	Alto: un positivo tiende a confirmar la enfermedad ^{F1,F4}	Disminuye: la OMS lo expresa al revés, cuanto más baja es la prevalencia, más probable es que un negativo no tenga la enfermedad ^{F5}

Ejemplo de Fisterra con un test de VIH de sensibilidad y especificidad aproximadas del 99,5 %, aplicado a los 2.800.000 habitantes de Galicia ^{F1}:

- Con 6.000 infectados (prevalencia del **0,21 %**), el test sería positivo en 19.940 personas y el **VPP sería del 29,9 %**: el 70,1 % de los positivos no tendrían la enfermedad ^{F1}.
- Con 800.000 infectados (prevalencia del **28,6 %**), el **VPP subiría al 98,7 %** y los falsos positivos bajarían a solo el 1,3 % de los positivos ^{F1}.
- Conclusión de los autores: si la prevalencia es alta, un resultado positivo tiende a confirmar la enfermedad; si es baja, un positivo no permite afirmar su existencia ^{F1}.



Lo que cambia y lo que no

Al pasar a una población de **mayor prevalencia aumenta el valor predictivo positivo** (y baja el negativo) ^{F1,F4}. **La sensibilidad y la especificidad no cambian**: son propiedades intrínsecas de la prueba e independientes de la prevalencia ^{F1,F4}. Tampoco cambian los cocientes de probabilidad (sección 6) ^{F1}.

La pandemia de COVID-19 lo hizo visible. La OMS insistía en que la prevalencia de la enfermedad en la comunidad estudiada **afecta fuertemente al valor predictivo** de un resultado positivo o negativo, y que la prevalencia debía estimarse con la vigilancia porque influye en el VPP y el VPN ^{F5}. En general, cuanto mayor es la prevalencia de infección en la población estudiada, más probable es que quien da positivo tenga COVID-19 ^{F5}.

- Con una prevalencia de infección activa del **1 %**, incluso una prueba con un 99 % de especificidad tendría un VPP pobre: **la mitad de los positivos serían falsos** ^{F5}.
- Con una prevalencia del **0,1 %**, una prueba con un 98 % de especificidad tendría un **VPP del 4 %**: 96 de cada 100 positivos serían falsos ^{F5}.
- Por eso la OMS consideraba las pruebas rápidas de antígenos más fiables donde la positividad era $\geq 5 \%$, y con transmisión nula o baja su VPP sería bajo (muchos falsos positivos) ^{F5}.
- El Ministerio de Sanidad, en su estrategia de diciembre de 2021, daba para una población con prevalencia de infección de entre el **10 % y el 30 %** (como las personas con síntomas en los entornos asistenciales) y

prueba hecha en los primeros 5 días, un VPN de entre el 97,2 % y el 99,3 % y un VPP de entre el 94,5 % y el 98,5 % ^{F6}.

De ahí se deduce lo que ocurriría entre las olas epidémicas: si la prevalencia sube y baja, lo que varía con la situación son los **valores predictivos**, porque dependen de la prevalencia ^{F5}; la sensibilidad y la especificidad, propiedades intrínsecas, no dependen de ella ^{F1}.

La prevalencia no mejora la prueba

Aplicar una prueba en un contexto de alta prevalencia **no la hace más sensible ni más específica** ni más fiable: cambia la probabilidad de que un positivo sea verdadero, es decir, el **valor predictivo positivo** ^{F1,F3}.

EIR 2019 · 210

P10 · respuesta en el solucionario

Valores predictivos: dependen de la prevalencia (alta prevalencia aumenta VPP)

Respecto a las pruebas diagnósticas, cuando se aplican en un contexto de alta prevalencia de la enfermedad, fundamentalmente debemos considerar que:

1. Aumenta la sensibilidad de la prueba.
2. Aumenta la especificidad de la prueba.
3. Aumenta el valor predictivo positivo de la prueba.
4. Aumenta el valor predictivo negativo de la prueba.

EIR 2022 · 50

P5 · respuesta en el solucionario

Valores predictivos: dependen de la prevalencia (alta prevalencia aumenta VPP)

Durante las olas de la pandemia por COVID19, en relación a las pruebas diagnósticas, ¿qué aspecto de los siguientes se verá principalmente influido por la situación?:

1. La sensibilidad.
2. La especificidad.
3. Los valores predictivos.
4. La fiabilidad.

6. Cocientes de probabilidad e índice de Youden

Como los valores predictivos dependen de la prevalencia, **no sirven para comparar dos pruebas** ni para extrapolar los resultados de otros estudios. Hacen falta índices clínicamente útiles que **no dependan de la prevalencia**: la **razón de verosimilitudes**, razón de probabilidad o **cociente de probabilidades**, que mide cuánto más probable es un resultado concreto según la presencia o ausencia de enfermedad ^{F1,F4}.

Índice	Cálculo	Lectura
Cociente de probabilidades positivo (CP+)	Sensibilidad / (1 – especificidad) = fracción de verdaderos positivos / fracción de falsos positivos ^{F1}	Cuántas veces es más probable un resultado positivo en un enfermo que en un sano ^{F1}
Cociente de probabilidades negativo (CP–)	(1 – sensibilidad) / especificidad = fracción de falsos negativos / fracción de verdaderos negativos ^{F1}	Probabilidad de un negativo en presencia de enfermedad frente a su probabilidad en ausencia de ella ^{F1}
Índice de Youden	Sensibilidad + especificidad – 1 ^{F4}	Otro valor sintético de la validez; las medidas sintéticas dan la misma importancia a sensibilidad y especificidad y, según Sancho-Insenser y González-Castillo, son de utilidad limitada ^{F4}

En el ejemplo del tacto rectal el **CP+ fue de 3,20**: un tacto anormal es unas 3 veces más probable en un paciente con cáncer de próstata que en uno sin cáncer ^{F1}.

- Relacionan la sensibilidad y la especificidad **en un solo índice** ^{F1,F4}.
- Pueden calcularse para **varios niveles** de un resultado, sin reducirlo a positivo/negativo ^{F1,F4}.
- Al igual que la sensibilidad y la especificidad, **no varían con la prevalencia**, por lo que sirven para **comparar pruebas** para un mismo diagnóstico ^{F1,F4}.

Qué depende de la prevalencia

Dependen: VPP y VPN ^{F1,F3}. **No dependen:** sensibilidad, especificidad, cocientes de probabilidad ^{F1,F4} y área bajo la curva ROC ^{F1}.

7. Curvas ROC y área bajo la curva

Cuando el resultado de la prueba es continuo (o una escala ordinal), cada valor de corte da un par de sensibilidad y especificidad. La **curva ROC** (*receiver operating characteristic*) representa gráficamente los pares **(1 – especificidad, sensibilidad)** obtenidos con **todos los posibles valores de corte** ^{F1,F2}. El análisis ROC nació en la teoría de la decisión, en los años 50, con la detección de señales por radar ^{F2}.

Elemento	Qué representa
Eje Y (ordenadas)	Sensibilidad = fracción de verdaderos positivos (FVP) ^{F2}
Eje X (abscisas)	1 – especificidad = fracción de falsos positivos (FFP) ^{F2}
Diagonal del vértice inferior izquierdo al superior derecho	Prueba que no discrimina entre sanos y enfermos ^{F2} ; equivale a un área de 0,5, como lanzar una moneda al aire ^{F4}
Vértice superior izquierdo	Discriminación perfecta (100 % de sensibilidad y 100 % de especificidad): la curva pasaría por ese punto ^{F2}
Desplazamiento de la curva	La exactitud aumenta a medida que la curva se aleja de la diagonal hacia el vértice superior izquierdo ^{F2}

La curva ROC es **necesariamente creciente**, reflejo del compromiso entre sensibilidad y especificidad: ganar sensibilidad moviendo el corte solo se consigue perdiendo especificidad ^{F2}. Sirve para comparar pruebas y para comparar varios **puntos de corte** de una misma prueba ^{F4}, y la elección del valor de corte se hace sobre ella teniendo en cuenta el coste de los falsos positivos frente a los falsos negativos y la prevalencia ^{F2}.



El área bajo la curva (ABC, AUC)

Es el **índice de la exactitud global** de la prueba: **1 = exactitud máxima** y **0,5 = mínima** (si saliera menor de 0,5, habría que invertir el criterio de positividad) ^{F2}. Fistera la considera **el mejor indicador de la capacidad predictiva del test, independiente de la prevalencia**, y la base para comparar pruebas diagnósticas ^{F1}: a mayor área, **mayor capacidad de discriminar correctamente entre enfermos y no enfermos** ^{F4}.

En términos probabilísticos, el área bajo la curva es **la probabilidad de que, si se eligen al azar un paciente enfermo y otro sano, el valor de la prueba sea mayor en el enfermo** ^{F2}. Se calcula con intervalo de confianza; con los datos de volumen corpuscular medio para diagnosticar anemia ferropénica, el ABC fue de 0,699 (IC 95 %: 0,585-0,813) ^{F2}. Al comparar dos pruebas se prefiere, en principio, la de mayor ABC, aunque dos curvas de forma muy distinta pueden tener un área casi igual y hay que mirarlas; para esos casos se proponen índices de **área parcial** ^{F2}.

¿El área va de 0 a 1 o de 0,5 a 1?

Matemáticamente el área oscila entre 0 y 1 ^{F4}, pero **0,5 corresponde al azar** (la diagonal) y un valor menor de 0,5 se corrige invirtiendo el criterio de positividad; por eso la exactitud útil va de **0,5 a 1** ^{F2}. Sancho-Insenser y González-Castillo aconsejan restar mentalmente 0,5: un área de 0,750 solo mejora en un 25 % el resultado de lanzar una moneda ^{F4}.

No confundir las herramientas: la **capacidad discriminativa** de una prueba (sanos frente a enfermos, en todos los puntos de corte) se resume con el **área bajo la curva ROC** ^{F1,F4}. El **índice kappa** mide otra cosa: la **concordancia** entre observadores corregida por la que se debe exclusivamente al azar ^{F9}. El **VPP** depende de la prevalencia ^{F1} y la **sensibilidad** aislada es un solo punto de la curva, que cambia con el corte elegido ^{F2}.

EIR 2020 · 70

P8 · respuesta en el solucionario

Curva ROC y área bajo la curva: capacidad discriminativa

Respecto al análisis de las curvas ROC de las pruebas diagnósticas, ¿qué afirmación de las siguientes es FALSA?:

1. El eje X representa los resultados falsos positivos.
2. El eje Y representa la sensibilidad.
3. Una prueba con una discriminación perfecta sanos-enfermos dará lugar a una recta en diagonal de 45°.
4. Permiten elegir el punto de corte apropiado para considerar resultados positivos y negativos de una prueba.

EIR 2025 · 36

P2 · respuesta en el solucionario

Curva ROC y área bajo la curva: capacidad discriminativa

Para evaluar la capacidad discriminativa (sanos vs enfermos) de un test o prueba diagnóstica, ¿cuál de las siguientes métricas utilizaría?:

1. El Área Bajo la Curva.
2. El índice Kappa.
3. El Valor Predictivo Positivo.
4. La Sensibilidad.

8. Pruebas en serie y en paralelo, y calidad de los estudios

Cuando se usan varias pruebas complementarias, se llaman **en paralelo** si se hacen simultáneamente y **en serie** si cada una se decide según los resultados de las anteriores ^{F1}.

Estrategia	Qué se gana	Qué se arriesga
En paralelo (a la vez)	Aumenta la probabilidad de diagnosticar a un enfermo ^{F1}	Aumenta también la probabilidad de considerar enfermo a un sano ^{F1}
En serie (una tras otra)	Pocos sanos serán considerados enfermos ^{F1}	El riesgo es no diagnosticar a algunos enfermos ^{F1}

Relacionado con ello: las pruebas de cribado deben ser de alta sensibilidad ^{F1}, y antes de iniciar un tratamiento se suele pedir una prueba confirmatoria, idealmente de alta especificidad ^{F4}.

Los **estudios de evaluación de pruebas diagnósticas** son el instrumento para obtener la sensibilidad, la especificidad y los valores predictivos ^{F1}. Lo que se obtiene son **estimaciones** para una población teórica, por lo que deben acompañarse de su precisión, idealmente con **intervalos de confianza** ^{F1,F2,F4}. La población de estudio, la estrategia de muestreo, la selección del **criterio de referencia** y la forma de aplicar las pruebas son elementos que hay que cuidar para evitar **sesgos** ^{F1}.



Cada parámetro, para quién

Sensibilidad y especificidad describen bien la validez de una prueba (ideales para el epidemiólogo) pero exigen conocer el diagnóstico real; los valores predictivos no lo exigen (ideales para el clínico) pero son muy sensibles a valores extremos de la prevalencia; los cocientes de probabilidad son independientes de la prevalencia (ideales para el investigador); el valor global tiene un valor muy limitado ^{F4}.

Todas las preguntas del tema (11)

De la más reciente a la más antigua. Las respuestas, en el solucionario.

EIR 2026 · 48

P1 · respuesta en el solucionario

En un estudio realizado en atención primaria para validar un test rápido de detección de antígenos del estreptococo β -hemolítico del grupo A en pacientes con faringitis, el valor predictivo positivo del test fue del 90 %. ¿Qué interpretación es correcta?:

1. Que el test identifica correctamente al 90 % de los pacientes con infección estreptocócica.
2. Que el 90 % de los pacientes con un resultado positivo en el test realmente presentan faringitis estreptocócica.
3. Que el test ofrece un resultado negativo en el 90 % de los pacientes sin infección.
4. Que el test acierta el 90 % de las veces, tanto en positivos como en negativos.

EIR 2025 · 36

P2 · respuesta en el solucionario

Para evaluar la capacidad discriminativa (sanos vs enfermos) de un test o prueba diagnóstica, ¿cuál de las siguientes métricas utilizaría?:

1. El Área Bajo la Curva.
2. El índice Kappa.
3. El Valor Predictivo Positivo.
4. La Sensibilidad.

EIR 2024 · 77

P3 · respuesta en el solucionario

En una prueba de cribado, la proporción de verdaderos positivos entre los identificados como positivos, se denomina:

1. Prevalencia.
2. Especificidad.
3. Sensibilidad.
4. Valor predictivo.

EIR 2023 · 21

P4 · respuesta en el solucionario

Cuando se habla de valor global en el proceso de toma de decisiones diagnósticas, hace referencia a:

1. El número de enfermos que existe en una población determinada en un período de tiempo concreto.
2. Es el número total de positivos obtenidos en el proceso diagnóstico en relación al total de sujetos que componen la muestra estudiada.
3. La proporción de sujetos correctamente clasificados por el test en relación al total de sujetos que componen la muestra de estudio.
4. Grado en que los resultados de un estudio son generalizables a otros sujetos o poblaciones.

Durante las olas de la pandemia por COVID19, en relación a las pruebas diagnósticas, ¿qué aspecto de los siguientes se verá principalmente influido por la situación?:

1. La sensibilidad.
2. La especificidad.
3. Los valores predictivos.
4. La fiabilidad.

Si sospechamos la presencia de una enfermedad severa, cuyo manejo conlleva medidas invasivas con potenciales efectos secundarios graves, ¿cuál sería la principal característica que buscaremos en la prueba diagnóstica a realizar?:

1. Tener una elevada sensibilidad.
2. Tener una elevada especificidad.
3. Tener un elevado valor predictivo positivo.
4. Tener un elevado valor predictivo negativo.

La Asociación Americana de Diabetes en 1997 y la Organización Mundial de la Salud en 1998 aprobaron nuevos criterios para el diagnóstico de la Diabetes Mellitus. Se recomendó el descenso del valor de glucemia basal de 140 mg/dl a 126 mg/dl, para diagnosticar a una persona de Diabetes. Este cambio de criterio consigue:

1. Incrementar la sensibilidad de la prueba.
2. Incrementar la especificidad de la prueba.
3. Disminuir los falsos positivos de la prueba.
4. Incrementar la fiabilidad de la prueba.

Respecto al análisis de las curvas ROC de las pruebas diagnósticas, ¿qué afirmación de las siguientes es FALSA?:

1. El eje X representa los resultados falsos positivos.
2. El eje Y representa la sensibilidad.
3. Una prueba con una discriminación perfecta sanos-enfermos dará lugar a una recta en diagonal de 45°.
4. Permiten elegir el punto de corte apropiado para considerar resultados positivos y negativos de una prueba.

Se ha realizado un estudio de validación de una prueba diagnóstica, cuyos resultados se reflejan en una tabla de contingencia. ¿Cuál de las siguientes afirmaciones es la correcta?:

1. El valor global se calcula mediante la fórmula: $a / (a+b+c+d)$.
2. La especificidad se calcula mediante la fórmula: $c / (a+c)$.
3. La sensibilidad se calcula mediante la fórmula: $a / (a+c)$.
4. El Valor Predictivo Negativo se calcula mediante la fórmula: $(b+c) / (a+b+c+d)$.

Respecto a las pruebas diagnósticas, cuando se aplican en un contexto de alta prevalencia de la enfermedad, fundamentalmente debemos considerar que:

1. Aumenta la sensibilidad de la prueba.
2. Aumenta la especificidad de la prueba.
3. Aumenta el valor predictivo positivo de la prueba.
4. Aumenta el valor predictivo negativo de la prueba.

Al realizar una prueba diagnóstica, ¿cuáles de las siguientes opciones no es una característica probabilística?:

1. Sensibilidad.
2. Valor predictivo negativo.
3. Especificidad.
4. Límites patológicos.

Solucionario

N°	Pregunta	Respuesta oficial	Notas
P1	EIR 2026 · 48	2	
P2	EIR 2025 · 36	1	
P3	EIR 2024 · 77	4	
P4	EIR 2023 · 21	3	hoja del archivo oficial
P5	EIR 2022 · 50	3	
P6	EIR 2021 · 40	2	
P7	EIR 2021 · 39	1	
P8	EIR 2020 · 70	3	
P9	EIR 2020 · 68	3	
P10	EIR 2019 · 210	3	hoja del archivo oficial
P11	EIR 2016 · 207	4	hoja del archivo oficial

Fuentes

- **F1** Pita Fernández S, Pértegas Díaz S. Pruebas diagnósticas: sensibilidad y especificidad. Cad Aten Primaria. 2003;10:120-124. Fistera (Metodología de la investigación). Fecha de revisión: 07/12/2010. · <https://www.fistera.com/formacion/metodologia-investigacion/pruebas-diagnosticas-sensibilidad-especificidad/> · consultada 2026-09-27
- **F2** López de Ullibarri Galparsoro I, Pita Fernández S. Curvas ROC. Cad Aten Primaria. 1998;5(4):229-235. Fistera (Metodología de la investigación). Fecha de revisión: 25/09/2001. · <https://www.fistera.com/formacion/metodologia-investigacion/curvas-roc/> · consultada 2026-09-27
- **F3** Bonita R, Beaglehole R, Kjellström T. Epidemiología básica. 2.ª ed. Publicación Científica y Técnica n.º 629. Washington, D.C.: OPS; 2008. Capítulo 6, pp. 143-144 (prueba de detección y cuadro 6.5, Validez de una prueba de detección sistemática); capítulo 8, pp. 171-174 (anormalidad asociada con enfermedad; pruebas diagnósticas). Traducción al español de la reimpresión corregida de la segunda edición en inglés. · <https://iris.paho.org/handle/10665.2/3153> · consultada 2026-09-27
- **F4** Sancho-Insenser JJ, González-Castillo AM. Pruebas diagnósticas. ¿Cómo describir su validez? Cir Esp. 2022;100(9):590-594 (carta metodológica, tablas 1 y 2). doi:10.1016/j.ciresp.2022.02.006. · <https://doi.org/10.1016/j.ciresp.2022.02.006> · consultada 2026-09-27
- **F5** Organización Mundial de la Salud. Antigen-detection in the diagnosis of SARS-CoV-2 infection: interim guidance, 6 October 2021 (WHO/2019-nCoV/Antigen_Detection/2021.1). Puntos clave (p. 1), nota 1 (p. 3), p. 4 y anexo (p. 17, tabla 1). · https://www.who.int/publications/i/item/WHO2019-nCoVAntigen_Detection2021.1 · consultada 2026-09-27
- **F6** Ministerio de Sanidad. Estrategia de detección precoz, vigilancia y control de COVID-19. Actualizado a 22 de diciembre de 2021 (aprobado por la Ponencia de Alertas y la Comisión de Salud Pública del Consejo Interterritorial). Nota 6, p. 10. Documento histórico de la fase aguda de la pandemia, sustituido en marzo de 2022 por la estrategia de vigilancia y control tras la fase aguda. · https://www.sanidad.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov/documentos/COVID19_Estrategia_vigilancia_y_control_e_indicadores.pdf · consultada 2026-09-27
- **F7** Alberti KG, Zimmet PZ. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus. Provisional report of a WHO consultation. Diabet Med. 1998;15(7):539-553 (resumen, leído en PubMed mediante las E-utilities del NCBI; PMID 9686693). · [https://doi.org/10.1002/\(SICI\)1096-9136\(199807\)15:7<539::AID-DIA668>3.0.CO;2-S](https://doi.org/10.1002/(SICI)1096-9136(199807)15:7<539::AID-DIA668>3.0.CO;2-S) · consultada 2026-09-27
- **F8** Organización Mundial de la Salud y Federación Internacional de Diabetes. Definition and diagnosis of diabetes mellitus and intermediate hyperglycaemia: report of a WHO/IDF consultation. Ginebra: OMS; 2006 (ISBN 92 4 159493 4). Resumen y recomendaciones (pp. 1-3) y pp. 22-23 (estudio ARIC). · <https://www.who.int/publications/i/item/definition-and-diagnosis-of-diabetes-mellitus-and-intermediate-hyperglycaemia> · consultada 2026-09-27
- **F9** López de Ullibarri Galparsoro I, Pita Fernández S. Medidas de concordancia: el índice kappa. Cad Aten Primaria. 1999;6:169-171. Fistera (Metodología de la investigación). Fecha de revisión: 01/12/2010. · <https://www.fistera.com/formacion/metodologia-investigacion/medidas-concordancia-indice-kappa/> · consultada 2026-09-27

Pendiente de verificar

- Valor previo de la glucemia basal diagnóstica de diabetes (140 mg/dl, 7,8 mmol/l) antes del cambio de 1997-1998: no confirmado en un texto oficial legible (el informe OMS de 1999 en IRIS es una imagen escaneada sin texto y el informe de la ADA de 1997 en Diabetes Care está tras una verificación contra robots). El tema

solo afirma el nuevo valor ($\geq 7,0$ mmol/l, 126 mg/dl). Buscar una edición en texto del informe OMS 1999 (WHO/NCD/NCS/99.2).