



Revisión de artículo

“GUÍA SOBRE PRECISIÓN Y EXACTITUD DEL EXAMEN CLÍNICO”

Traducción adaptada del artículo “A Primer on the Precision and Accuracy of the Clinical Examination”, del doctor David L. Sackett, MD, MSc Epid, FRCPC (Department of Medicine, Faculty of Medicine, McMaster University, Hamilton, Ontario, Canada), publicado en la revista JAMA (1192; 267(19):2638-2644)

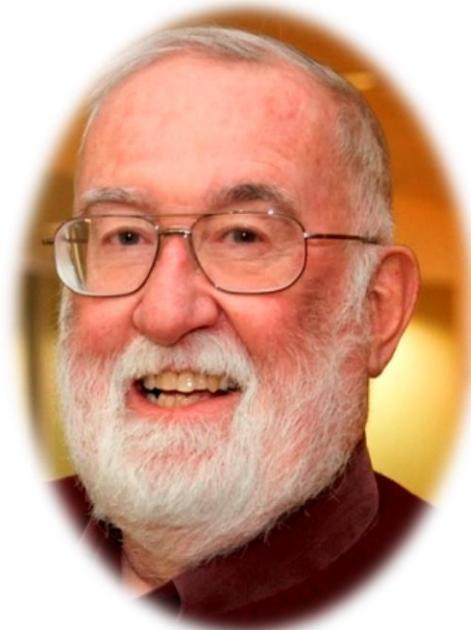
Disponible en:

http://66.199.228.237/boundary/complex_decision_making_and_ethics/A_primer_on_precision_accuracy_of_physical_exam.pdf



Jenny A. Pinzón****

Hemos traducido este inmejorable artículo en honor al doctor David Sackett (incorporándole algunas imágenes y resumiendo algunas partes), considerándolo muy importante por las herramientas que ofrece para el examen clínico racional y para evaluar artículos que traten sobre test de diagnóstico.



http://cdmmedhall.org/sites/default/files/sackett_final.jpg

Este artículo es un manual básico para el examen clínico racional que fue publicado en la revista JAMA como parte de una serie de revisiones relacionadas. Presenta definiciones y explicaciones de términos y conceptos claves; incluye una guía para evaluar artículos sobre test de diagnóstico, así como una tabla para desarrollar ejercicios. Las guías posteriores de la serie discuten temas seleccionados sobre la precisión y exactitud del examen médico con más detalle o los amplían para situaciones de más complejo diagnóstico.

**** Coeditora Carta Comunitaria, jenny-pinzon@juanncorpas.edu.co.



Si bien el artículo aclara que la precisión y exactitud en el examen clínico no son la única preocupación durante una revista médica, su aplicación ofrece un punto de partida para decidir cuánta certeza se necesita tener en un diagnóstico antes de intervenir sobre él.

EL PACIENTE



El autor presenta a un paciente para ejemplificar lo mencionado anteriormente:

Uno de sus pacientes, a quien no ha visto por muchos años, es ingresado al servicio de ortopedia después de que un cajón de embalaje se volcó sobre su pierna, produciendo una fractura inestable en su tibia distal y su peroné. Usted se detiene para verlo, mientras se le está preparando para una cirugía. Él está despierto y hemodinámicamente estable, pero huele a alcohol (a las 10:00 a.m.); tiene tres telangiectasias en la parte superior de su pecho (no presenta ginecomastia ni asterixis). Es obeso y su vientre es prominente. Dentro de las preguntas que surgen en su mente, las siguientes son las más significativas:

1. *¿Es alcohólico?* Usted podría asignar un 50-50 de probabilidad para este desorden (este cálculo se explica en otra guía de esta serie de la revista). La respuesta a esta pregunta de diagnóstico es importante tanto a largo plazo como para protegerlo de las complicaciones de una abstinencia aguda durante y después de la operación.

2. *¿Tiene ascitis?* Usted está mucho menos seguro de esto, pero si el paciente es dependiente del alcohol, podría decir que la probabilidad de que la prominencia de su vientre se deba a una ascitis es de un 50-50. De nuevo, sería importante saber si esta manifestación de daño avanzado del hígado se debe al alcohol.

Las opciones para responder estas preguntas son variadas. Para explorar posible abuso o dependencia de alcohol usted podría:

- a) Tomar el tiempo requerido para interrogarlo sobre la cantidad de alcohol que consume (distanciarse del personal de enfermería y, quizá, exasperarse por lo que implica este proceso).
- b) Ordenar una o más pruebas de función hepática.
- c) Solicitar un test de actividad enzimática plaquetaria, la cual se reporta elevada en alcohólicos.

d) Hacerle las cuatro preguntas rápidas del test de "CAGE" para detectar el alcoholismo (acrónimo en inglés para las preguntas: ¿Ha sentido alguna vez que debe beber menos? ¿Le ha molestado que la gente lo critique por su forma de beber? ¿Alguna vez se ha sentido mal o culpable por su forma de beber? ¿Alguna vez ha necesitado beber por la mañana para calmar los nervios o eliminar molestias por haber bebido la noche anterior?

Para explorar una posible ascitis, pudiera:

- a) Percutir para detectar matidez cambiante, onda líquida o incluso el signo del charco.
- b) Ordenar un examen de ultrasonido abdominal.
- c) Simplemente preguntarle si alguna vez ha tenido los tobillos inflamados.

Consideraciones



El autor invita a reflexionar en las implicaciones, en términos de tiempo y dinero, de las formas alternativas para responder las dos preguntas planteadas anteriormente (si el paciente abusa o es dependiente del alcohol y si presenta ascitis) y plantea la siguiente cuestión:

“¿No sería genial que usted pudiera responderlas con solos cinco preguntas rápidas (las cuatro del test de CAGE y la referente a la inflamación del tobillo)?”.

Y responde que, hasta cierto punto, esto se podría hacer pues si el paciente responde afirmativamente a tres o cuatro preguntas del test de CAGE, indicaría que él abusa o es dependiente del alcohol (una información mucho más poderosa que cualquier prueba de laboratorio que se pueda ordenar). Además, si el paciente responde que no se le han

inflamado los tobillos, se descartaría clínicamente una ascitis importante (se podría chequear nuevamente con el signo de matidez cambiante y notar que, como muchos de tales pacientes, no tiene la onda ascítica y que el signo del charco no es útil en estos casos). Así, para ambas preguntas un rápido examen ha arrojado información diagnóstica definitiva, sin la necesidad de una prueba de laboratorio o de una imagen diagnóstica.

Ahora plantea la cuestión de por qué se puede hacer esta osada declaración acerca del poder de estos simples elementos de la historia clínica y del examen físico. La respuesta radica en la ciencia del diagnóstico basado en el examen clínico racional.

TOMANDO UNA HISTORIA CLÍNICA ALTERNATIVA PARA EL ALCOHOLISMO

| | | Alcohol Abuse or Dependency | | |
|---|---------------|-----------------------------|-----------------|-----------|
| | | Yes | No | |
| No. of Positive Answers to the 4 CAGE Questions | 3 or 4 | 60 (True +) | 1 (False +) | 61 |
| | | a | b | $a+b$ |
| | | c | d | $c+d$ |
| | 2, 1, or None | 57 (False -) | 400 (True -) | 457 |
| | | $a+c$ | $b+d$ | $a+b+c+d$ |
| | | 117 | 401 | 518 |

Figura 1. Abuso o dependencia de alcohol de acuerdo al número de respuestas positivas al test de CAGE.



| Características: | |
|--|---|
| $\frac{a}{(a+c)} = \frac{60}{117} = 0,51 = 51\%$ | Sensibilidad |
| $\frac{d}{(b+d)} = \frac{400}{401} = 0,998 = 99,8\%$ | Especificidad |
| Predicciones: | |
| $\frac{a}{(a+b)} = \frac{60}{61} = 0,98 = 98\%$ | Valor predictivo positivo = Probabilidad postest de presentar el desorden estudiado (abuso o dependencia del alcohol) en pacientes con 3 o 4 respuestas positivas. |
| $\frac{d}{(c+d)} = \frac{400}{457} = 0,88 = 88\%$ | Valor predictivo negativo = Probabilidad postest de NO presentar el desorden estudiado (abuso o dependencia del alcohol) en pacientes con 2 o menos respuestas positivas. |
| $\frac{c}{(c+d)} = \frac{57}{457} = 0,12 = 12\%$ | = Probabilidad postest de presentar el desorden estudiado (abuso o dependencia del alcohol) en pacientes con 2 o menos respuestas positivas. |
| Probabilidad pretest (prevalencia) | |
| $\frac{(a+c)}{(a+b+c+d)} = \frac{117}{518} = 23\%$ | =Prevalencia o probabilidad pretest de tener el desorden estudiado. |

La figura 1 muestra el número total de respuestas positivas al test de CAGE de dos grupos diferentes de pacientes admitidos en ortopedia y en los servicios médicos de un hospital de enseñanza comunitario en Boston (Massachusetts). En la columna de la izquierda, están las respuestas de pacientes cuyo chequeo exhaustivo (incluyendo, donde era indicado, historias sociales detalladas, seguimientos y biopsias del hígado) arrojó "prueba" aceptable de que tenían problema con el alcohol (abusaban o eran dependientes del alcohol). En la columna de la derecha están los pacientes cuyo chequeo mostró que no tenían dicho problema.

¿Cómo determinar si son válidos los resultados de un estudio sobre pruebas diagnósticas?

Tabla 1. Guía para leer artículos sobre pruebas diagnósticas

| |
|--|
| 1. ¿Se hizo una comparación independiente, "ciega", con una prueba de diagnóstico "gold estándar"? |
| 2. ¿Se evaluó el test de diagnóstico en un espectro adecuado de pacientes (que incluya sujetos con enfermedad leve y severa, tratada y no tratada e individuos con enfermedades diferentes, pero con un cuadro clínico similar)? |
| 3. ¿Se describieron adecuadamente tanto el contexto como la selección de los pacientes? |
| 4. ¿Se determinó la forma de replicar el resultado del test (precisión) y su interpretación? |
| 5. ¿Se definió, de forma razonable, cómo se aplica el término "normal" en el test? |
| 6. Si se propusiera el test como parte de un grupo o secuencia de varios test, ¿se ha determinado su contribución individual en la validación general de estos? |
| 7. ¿Se han descrito con suficiente detalle las tácticas para realizar el test como para replicarlo exactamente? |
| 8. ¿Se ha determinado la utilidad de la prueba? |

En este punto el artículo aclara que este tipo de investigaciones muchas veces son consultadas como pruebas "gold estándar" del diagnóstico y típicamente consisten en hallazgos definitivos de angiografías, procedimientos, autopsias y similares.

En el caso del estudio sobre el test de CAGE, al examinarlo a la luz de la "Guía para leer artículos sobre pruebas diagnósticas" (ver tabla 1), la cual establece unos criterios para determinar si un artículo sobre estrategias diagnósticas es válido y clínicamente útil, se puede concluir lo siguiente:

- ✓ El estudio se realizó de forma independiente y dentro de un espectro de pacientes bien descritos en quienes fue clínicamente razonable preguntar sobre problemas con el consumo de alcohol (criterio 1).
- ✓ El estudio se realizó en una muestra de pacientes que incluía pacientes con el desorden estudiado (abuso o dependencia del alcohol) leve y severo, tratados y no tratados y con enfermedades diferentes pero con un cuadro clínico similar (criterio 2).
- ✓ El contexto del estudio (un gran hospital general urbano) fue descrito de tal forma que permite determinar la aplicabilidad de los resultados en otros contextos (criterio 3).
- ✓ El criterio 4 respecto a la replicabilidad de los resultados del estudio se analiza en otra revisión.

- ✓ La clasificación de un resultado "normal" en el test fue claramente definida como la ausencia de abuso o dependencia del alcohol (criterio 5).
- ✓ Los autores del estudio CAGE no propusieron que sus preguntas se usaran como parte de una amplia serie de pruebas de diagnóstico (no aplica el criterio 6).
- ✓ Las preguntas del test de CAGE se redactaron y presentaron exactamente en el artículo permitiendo así su aplicación exacta en la práctica (criterio 7).
- ✓ El punto final de la guía para el lector (*¿se ha determinado la utilidad de la prueba?*) se satisface en la medida en que las preguntas del CAGE reconocen al alcohólico, especialmente a quienes abusan del alcohol, mucho más que el diagnóstico clínico de rutina, haciéndolos candidatos para tratamiento y terapia.

En resumen, el estudio CAGE sigue los estándares metodológicos requeridos para una válida y útil descripción sobre la aplicabilidad clínica de cualquier información diagnóstica, tanto si viene de la historia clínica, del examen físico o de un diagnóstico de laboratorio.

LA PRECISIÓN DEL EXAMEN CLÍNICO

Para que un artículo sobre historia clínica o examen clínico se considere exacto, debe primero ser preciso. Lo anterior significa que se necesita tener un grado de confianza en el que dos médicos clínicos que examinan al mismo paciente coincidirán en la presencia o ausencia del síntoma (por ejemplo, la respuesta a una de las preguntas del test de

CAGE) o del signo (tal como la presencia de telangiectasias en el pecho del paciente descrito anteriormente). La precisión (que muchas veces aparece bajo el nombre "variación del observador" en la literatura clínica) de tales hallazgos clínicos puede ser cuantificada.

| | | First Clinician's Examination for Spider Nevi | | |
|--|----------|---|-------------------|---------|
| | | Positive | Negative | |
| Second Clinician's Examination for Spider Nevi | Positive | 23 (Expect 8) | 5 | 28 |
| | Negative | 6 | 66 (Expect 51) | 72 |
| | | a+c | b+d | a+b+c+d |
| | | 29 | 71 | 100 |

Figura 2. Precisión del examen clínico para telangiectasias.

| | |
|--|---|
| $\frac{(a + d)}{(a + b + c + d)} = \frac{(23 + 66)}{100} = 89\%$ | Concordancia observada. |
| $\frac{([a + b] \times [a + c])}{(a + b + c + d)} = \frac{(28 \times 29)}{100} = 8$ | Concordancia observada para la celda a. |
| $\frac{([c + d] \times [b + d])}{(a + b + c + d)} = \frac{(72 \times 71)}{100} = 51$ | Concordancia observada para la celda d. |
| $\frac{(\text{Esperada } a - \text{esperada } d)}{(a + b + c + d)} = \frac{(8 + 51)}{100} = 59\%$ | |
| $\frac{(\text{Acuerdo observado} - \text{acuerdo esperado})}{(100\% - \text{acuerdo esperado})} = \frac{(89\% - 59\%)}{(100\% - 59\%)} = 0,73$ | Concordancia más allá de la casualidad = κ |
| Niveles de concordancia convencionales para κ: | |
| 0,0 - 0,2 | Leve. |
| 0,2 - 0,4 | Adecuada. |
| 0,4 - 0,6 | Moderada. |
| 0,6 - 0,8 | Significativa. |
| 0,8 - 1,0 | Casi perfecta. |



Por ejemplo, dos clínicos examinan, de manera independiente, los mismos 100 pacientes sospechosos de padecer enfermedad hepática y registran los siguientes datos (figura 2) sobre el hallazgo de telangiectasias:

- Los dos clínicos concuerdan en que 23 de los pacientes (celda a) tienen telangiectasias y en que 66 pacientes (celda d) no las presentan; así ellos coinciden en $(23+66)/100=89\%$ de los que examinaron.
- Sin embargo, a diferencia del segundo clínico, el primero afirmó que 6 pacientes (celda c) presentaban telangiectasias y que cinco pacientes (celda b) no las presentaban.

¿Cómo se debería interpretar esa precisión?

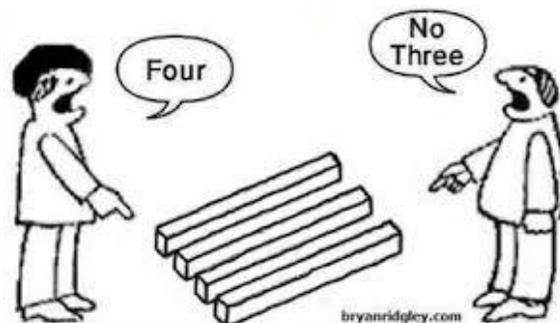


Imagen tomada de: https://cdn-images-1.medium.com/max/1600/1*0nyTsAqo-0Dn-Son33HA.jpeg

¿Es este nivel de concordancia clínica bueno o se debe esperar uno mejor? Para contestar esta pregunta se puede empezar por reconocer que algo de la concordancia clínica pudiera ocurrir por casualidad. Por ejemplo, suponiendo que el segundo clínico lanzó una moneda por cada paciente en

vez de realizar el examen y reportó la presencia de telangiectasias cada vez que la moneda cayó por el lado de la cara y, lo contrario, la no presencia del signo si cayó por el lado del sello, se tendría una concordancia del 50 %.

Concordancia no debida al azar

Se debe comenzar, entonces, por determinar cuánto de la concordancia observada (89 %) se debió al azar y así descubrir cuánta habilidad clínica mostraron los dos clínicos (acuerdo no debido al azar).

La concordancia debida al azar se puede calcular a través de un proceso formal, pero se podría pensar en un lanzamiento de moneda, en el cual, por ejemplo:

→ La moneda del primer clínico salió "cara" el 29 % de las veces (basado en $[a+c]/[a+b+c+d]$). Así, el 29 %, de los 28 pacientes evaluados con telangiectasias por el segundo clínico (a+b) serían diagnosticados con estas por

el primero y el 29 % de 28 es 8 (el número de pacientes que esperaríamos encontrar en la celda a solo por azar).

→ De manera similar, la moneda del primer clínico salió "sello" 71 % de las veces ($[b+d]/[a+b+c+d]$), y el 71 % de los 72 pacientes que no tienen telangiectasias, de acuerdo con el segundo clínico (c+d) corresponde a 51 (el valor esperado en la celda d).

→ Como resultado, se esperaría que los dos clínicos estuvieran de acuerdo $(8+51)/100$ o 59 % de las veces basados solo en el azar y el potencial remanente de concordancia más allá del azar es, por lo tanto, 100 % - 59 %, o 41 %.

¿Cuánto de este potencial de concordancia (41 %) no debido al azar se alcanzó?

Eso se determina al compararlo con el actual potencial de acuerdo más allá del azar, es decir:

$$89\% - 59\% = 30\%$$

Y,

$$30\% \div 41\% = 0,73$$

Lo cual significa que los clínicos alcanzaron cerca de tres cuartos de concordancia potencial no debida al azar.

Esta medida de concordancia se conoce con el nombre de *kappa* (κ) y es como un coeficiente de correlación cuyo rango va de -1,0 (donde dos clínicos estarían en perfecto desacuerdo) pasando por 0,0 (donde solo se alcanza la concordancia por el azar), hasta +1,0 (donde los dos clínicos estarían en perfecta concordancia).

De acuerdo con la lista de "niveles convencionales para κ " que aparece en el epígrafe de la figura 2, la concordancia de los dos clínicos se considera "significativa" y este es el caso para muchos hallazgos "presentes/ausentes" del examen médico. Como es de esperar, esta medida de concordancia

es mayor aun cuando los exámenes clínicos los realiza el mismo clínico.

A otros aspectos del examen médico no les va tan bien. Por ejemplo, en el estudio del examen del torso, los coeficientes κ para cianosis, taquipnea y pectoriloquio susurrado, fueron 0,36, 0,25 y, 0,11, respectivamente.

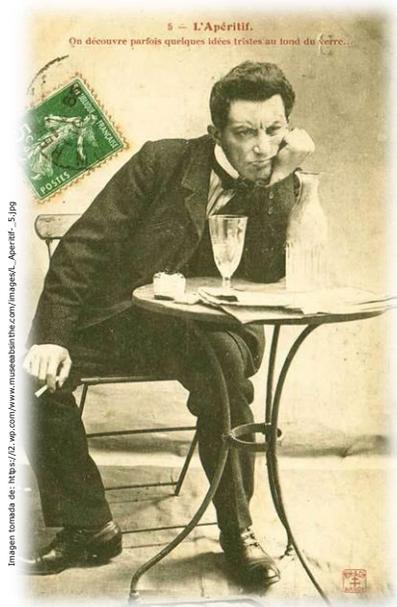
Ninguna medida para la concordancia clínica es ideal y el coeficiente κ no es la excepción. Su magnitud se ve afectada ligeramente por la frecuencia de un hallazgo anormal en el grupo de pacientes examinados (esta es más alta cuando la mitad de los pacientes tienen el hallazgo y disminuye un poco cuando el hallazgo es extremadamente común o raro).

La precisión no es suficiente

El artículo pasa ahora a abordar el tema de la "exactitud", afirmando que aunque la precisión sea consistente, puede ser inapropiada para las estimaciones de los investigadores. Menciona como ejemplo el que, ocasionalmente, los clínicos fallan en detectar un hígado aumentado o en escuchar un soplo diastólico. Igualmente cita el caso de un grupo de iridólogos que obtuvieron solo una concordancia clínica leve con un coeficiente κ promedio de 0,18, cuando se les solicitó que examinaran a una serie de pacientes para distinguir aquellos que tuvieran cálculos biliares de quienes no los

tenían de acuerdo al ultrasonido. Más relevante aun es que su exactitud diagnóstica no fue mejor que el azar pues fallaron en cerca de la mitad de los pacientes con cálculos biliares (sensibilidad=54 %) y diagnosticaron cálculos biliares en cerca de la mitad de pacientes con ultrasonido negativo (especificidad=52 %). Así, entonces, se hace necesario determinar las características de la exactitud, dejando a un lado la precisión, del examen clínico para entender la sensibilidad y la especificidad.

CARACTERÍSTICAS DE EXACTITUD DE LAS PRUEBAS DIAGNÓSTICAS



| | | Alcohol Abuse or Dependency | | |
|---|---------------|-----------------------------|-----------------|-----------|
| | | Yes | No | |
| No. of Positive Answers to the 4 CAGE Questions | 3 or 4 | 60 (True +) | 1 (False +) | 61 |
| | | a | b | $a+b$ |
| | | c | d | $c+d$ |
| | 2, 1, or None | 57 (False -) | 400 (True -) | 457 |
| | | $a+c$ | $b+d$ | $a+b+c+d$ |
| | | 117 | 401 | 518 |

Retomando la figura 1. . Abuso o dependencia de alcohol de acuerdo al número de respuestas positivas al test de CAGE.

Volviendo a la figura 1, podemos examinar las características de exactitud del test de CAGE:

60 pacientes de la celda *a*, respondieron "sí" a 3 o 4 preguntas del test y constituyen el 51 % (0,51) de los 117 pacientes (*a*+*c*) con diagnóstico positivo de abuso o dependencia de alcohol. El término clave para esta proporción de 0,51, o $a/(a+c)$, es **sensibilidad** y es una medida útil para determinar **qué tan bien** una prueba diagnóstica (sea un síntoma, un signo o prueba de laboratorio) **detecta** el desorden objeto **cuando está presente**. Cuanto más cerca esté la sensibilidad al 100 %, más "sensible" el hallazgo clínico o de laboratorio.

En la columna de la derecha están las respuestas de los pacientes en quienes la prueba descartó problemas con el alcohol. 400 pacientes en la celda *d*, respondieron "sí" a dos, una o a ninguna de las preguntas y constituyen el 99,8 % (0,998) del total de 401 pacientes (*b*+*d*) que no presentaban dependencia o abuso de alcohol. El término clave para esta proporción (0,998) o $d/(b+d)$ es **especificidad** y es útil para medir **con qué frecuencia un síntoma, signo o cualquier**

prueba de test está ausente cuando el desorden estudiado no está presente. Cuanto más cerca al 100 % esté la especificidad, más "específico" el hallazgo clínico o de laboratorio (por supuesto, el impacto de la sensibilidad y la especificidad en la interpretación de los hallazgos tanto positivos como negativos es un asunto que interesa mucho a los clínicos).

El hecho de que la sensibilidad del test sea solo del 51 % no es sorprendente dado que el número de "**verdaderos positivos**" en la celda *a* es casi igual al número de "**falsos negativos**" de la celda *c* y, dicho porcentaje confirma que la prueba falla en detectar el problema en cerca de la mitad del problema estudiado.

De otra parte, la especificidad de las preguntas del test de CAGE es espectacular. El número de los "**verdaderos negativos**" en la celda *d* supera enormemente la cantidad de "**falsos positivos**" de la celda *b* y una especificidad del 98 % confirma que casi nunca cataloga a un paciente de bebedor problemático cuando el desorden está ausente.

Los valores predictivos



El arte de predecir se basa en las características anteriores. En este caso se puede considerar que:

- Debido a la alta especificidad, prácticamente todos los pacientes de la celda *a* (la mayoría de los que respondieron afirmativamente a tres o cuatro preguntas del test, es decir, $a + b$) presentan el problema estudiado (abuso o dependencia del alcohol) y la probabilidad de que sean verdaderos positivos corresponde al "**valor predictivo positivo**", el cual se calcula de la siguiente manera:

$$\frac{a}{(a + b)} = \frac{60}{61} = 98 \%$$

Este valor predictivo positivo es lo mismo que la *probabilidad postest* de presentar el problema estudiado (entre los pacientes con tres o cuatro respuestas positivas).

- A pesar de que la sensibilidad es baja, la mayoría de los pacientes de las celdas *c* y *d* que respondieron afirmativamente a ninguna, una sola o a dos preguntas están en la celda *d* y no tenían el problema estudiado y, la probabilidad de que sean verdaderos negativos corresponde al "**valor predictivo negativo**", el cual se calcula de la siguiente manera:

$$\frac{d}{(c + d)} = \frac{400}{457} = 88 \%$$

Este valor predictivo negativo es lo mismo que la *probabilidad postest* de **no** presentar el problema estudiado (entre los pacientes con dos o menos respuestas positivas).

El complemento de este valor predictivo negativo, o $c/(c+d)$, describe la probabilidad de **tener** el desorden en aquellos pacientes con dos o menos respuestas positivas y algunos clínicos encuentran útil esta otra forma de decir lo mismo.

- La razón para que el valor predictivo negativo parezca relativamente alto, a pesar de la baja sensibilidad, se basa, para empezar, en el hecho de que la proporción de todos los pacientes en este estudio que presentaban abuso o dependencia del alcohol, $(a + c) / (a + b + c + d)$ o 117/518, eran solamente el 23 %. Esto significa que, 100 % - 23 % o, 77 % de los pacientes *no* eran dependientes del alcohol antes de responder cualquier pregunta del test. El término clave para el conocimiento previo contenido en esta expresión $(a + c)/(a + b + c + d)$ es "**prevalencia**" o, más útil, la "**probabilidad pretest**" de padecer el problema estudiado.

La prevalencia y los valores predictivos

En contraste con la probabilidad pretest de presentar el problema estudiado (23 %) en el artículo clínico que describe las preguntas del CAGE, aquí se ha juzgado que la probabilidad pretest de abuso o dependencia de alcohol es del 50 %. ¿Cómo se comportarían entonces las preguntas del test en pacientes como este?

| | | Alcohol Abuse or Dependency | | |
|---|---------------|-----------------------------|-----------------|----------------|
| | | Yes | No | |
| No. of Positive Answers to the 4 CAGE Questions | 3 or 4 | 510 (True +) | 2 (False +) | 512 |
| | 2, 1, or None | 490 (False -) | 998 (True -) | 1488 |
| | | <i>a</i> | <i>b</i> | <i>a+b</i> |
| | | <i>c</i> | <i>d</i> | <i>c+d</i> |
| | | <i>a+c</i> | <i>b+d</i> | <i>a+b+c+d</i> |
| | | 1000 | 1000 | 2000 |

Figura 3. Las preguntas del test de CAGE para abuso o dependencia del alcohol cuando la probabilidad pretest es del 50 %.

| Características: | |
|---|--|
| $\frac{a}{(a+c)} = \frac{510}{1000} = 0,51 = 51\%$ | Sensibilidad |
| $\frac{d}{(b+d)} = \frac{998}{1000} = 0,998 = 99,8\%$ | Especificidad |
| Probabilidad pretest (prevalencia) | |
| $\frac{(a+c)}{(a+b+c+d)} = \frac{1000}{2000} = 50\%$ | =Prevalencia o probabilidad pretest de tener el desorden estudiado. |
| Predicciones: | |
| $\frac{a}{(a+b)} = \frac{510}{512} = 0,996 = 99,6\%$ | Valor predictivo positivo=Probabilidad postest de presentar el desorden estudiado (abuso o dependencia del alcohol) en pacientes con 3 o 4 respuestas positivas. |
| $\frac{d}{(c+d)} = \frac{998}{1488} = 0,67 = 67\%$ | Valor predictivo negativo=Probabilidad postest de no presentar el desorden estudiado (abuso o dependencia del alcohol) en pacientes con 2 o menos respuestas positivas. |
| $\frac{c}{(c+d)} = \frac{490}{1488} = 0,33 = 33\%$ | =Probabilidad postest de tener el desorden estudiado (abuso o dependencia del alcohol) en pacientes con 2 o menos respuestas positivas. |

Siempre y cuando el conjunto de pacientes y la intensidad de la enfermedad en el estudio sobre el test CAGE sean similares a los que se presentan en la práctica del clínico que atiende al paciente descrito en este artículo, se espera que la sensibilidad y la especificidad permanezcan constantes, a pesar de la estimación de la probabilidad pretest de dicho paciente. Así, la sensibilidad (51 %) y la especificidad (99,8 %) en la figura 3 son las mismas que las de la figura 1.

Note, sin embargo, que el valor predictivo negativo ha descendido de 88 % a 67 %. Esto es debido a que los valores

predictivos **deben cambiar** con los **cambios en la prevalencia** del desorden estudiado. Después de todo el valor predictivo de un test positivo es simplemente la prevalencia de dicho desorden entre aquellos pacientes con resultados positivos en el test. Igualmente, el valor predictivo negativo es la prevalencia de **no tener** el desorden entre los pacientes con un resultado negativo del test. No sorprende entonces que los valores predictivos cambien con la variación de la prevalencia.

RETORNANDO AL PACIENTE



Imagen tomada de: <http://fotografias.antena3.com/clipping/cmsimagenes01/2015/12/14/AE7ECBE9-3357-4C4D-8747-1628A6E466E3/58.jpg>

El paciente rápidamente admite que ha reducido el consumo de alcohol, que su esposa y sus compañeros de trabajo lo han incomodado al quejarse por su manera de beber y que él muchas veces necesitaba un trago para comenzar el día. Basado en esta rápida historia y dando su primera apreciación sobre la probabilidad de ser dependiente del alcohol (50-50 o, lo que es lo mismo, una probabilidad pretest de 50 %), usted puede concluir, de acuerdo con la figura 3, que su probabilidad posttest de dependencia del alcohol es de 99,6 %.

Esto indica que cuando la **especificidad** es extremadamente **alta**, un test con **resultado positivo, diagnostica el desorden** estudiado (la especificidad del test es de 99,8 % de acuerdo a la figura 3). Compare lo anterior con los resultados, de la misma investigación del CAGE, para las especificidades del γ -glutamyl transpeptidase (76 %), del volumen corpuscular medio (64 %) y de un análisis completo de la función hepática (81 %). Además, el test para la actividad enzimática plaquetaria tiene una especificidad de solo 73 %. Así, en el paciente en cuestión, un sencillo historial no solo fue rápido y fácil de realizar sino que también fue más específico que los test de laboratorio.

¿Qué hay de la posible ascitis?

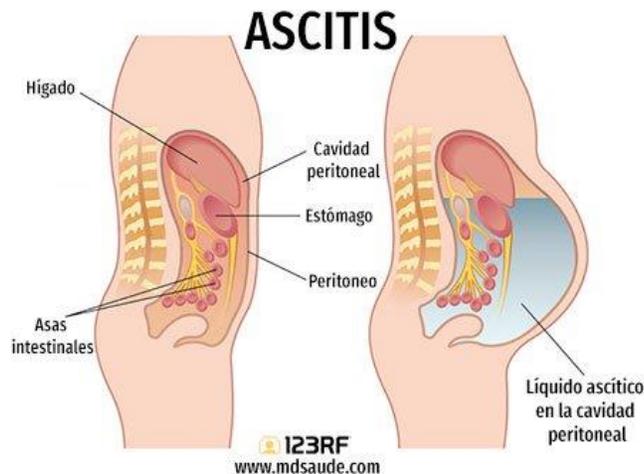


Imagen tomada de: <https://www.mdsau.de/wp-content/uploads/ascitis-ilustracion.jpg>

Dado que ya se ha establecido el diagnóstico de dependencia de alcohol, ya se puede planear el tratamiento pre y postoperatorio para prevenir, detectar y tratar los síntomas de la abstinencia de alcohol. No obstante, es necesario saber si el paciente tiene suficiente daño hepático como para interferir en el posible manejo de los medicamentos que se requieran. Para determinarlo, el clínico considera que, por la fractura de tobillo, la posición de rodillas requerida para provocar el signo del charco es imposible y que, aun un test de matidez desplazable, le provocaría un dolor considerable. Además, como el paciente ya ha estado en radiología sería mejor evitar otro desplazamiento allí para realizarle un ultrasonido abdominal. El tobillo que no está comprometido no está inflamado en este momento y él admite que antes nunca se le han inflamado los tobillos. ¿Servirá este historial?



| | | Presence Of Ascites on Abdominal Ultrasound | | |
|---------------------------|-----|---|------------|----------------|
| | | Present | Absent | |
| History of Ankle Swelling | Yes | 14 | 16 | 30 |
| | | <i>a</i> | <i>b</i> | <i>a+b</i> |
| No | | 1 | 32 | 33 |
| | | <i>c</i> | <i>d</i> | <i>c+d</i> |
| | | <i>a+c</i> | <i>b+d</i> | <i>a+b+c+d</i> |
| | | 15 | 48 | 63 |

Figura 4. Relación entre un historial de inflamación de tobillos y ascitis.

La figura 4 resume un estudio de 63 pacientes que ingresaron a un servicio médico general en Durham, Carolina del Norte. De 15 pacientes con ascitis determinada a partir de ultrasonido abdominal (gold estándar para este signo), 14 tienen historial de tobillos inflamados, con una sensibilidad del 93 %. Si se aplica esta sensibilidad (93 %) y la especificidad (66 %) al pretest de probabilidad para ascitis de 50 % en el caso del paciente estudiado en este artículo,

el resultado (que se muestra en la figura 5) sugiere que la probabilidad postest de no tener ascitis es del 90 % cuando el paciente niega inflamación en los tobillos. De nuevo, este sencillo dato de la historia clínica provee una poderosa información diagnóstica: cuando la *sensibilidad* de un síntoma o signo es alta, una respuesta negativa descarta el desorden estudiado

| Características: | |
|---|--|
| $\frac{a}{(a+c)} = \frac{14}{15} = 0,93 = 93\%$ | Sensibilidad |
| $\frac{d}{(b+d)} = \frac{32}{48} = 0,67 = 67\%$ | Especificidad |
| Predicciones: | |
| $\frac{a}{(a+b)} = \frac{14}{30} = 0,47 = 47\%$ | Valor predictivo positivo=Probabilidad postest de presentar el desorden estudiado (ascitis) en pacientes con historial de inflamación de tobillos. |
| $\frac{d}{(c+d)} = \frac{32}{33} = 0,97 = 97\%$ | Valor predictivo negativo=Probabilidad postest de no presentar el desorden estudiado (ascitis) en pacientes con historial negativo de inflamación de tobillos. |
| $\frac{c}{(c+d)} = \frac{1}{33} = 0,03 = 3\%$ | =Probabilidad postest de presentar el desorden estudiado (ascitis) en pacientes con historial negativo de inflamación de tobillos. |

| | | Presence of Ascites on Abdominal Ultrasound | | |
|---------------------------|-----|---|------------|----------------|
| | | Present | Absent | |
| History of Ankle Swelling | Yes | 93 | 34 | 127 |
| | | <i>a</i> | <i>b</i> | <i>a+b</i> |
| No | | 7 | 66 | 73 |
| | | <i>c</i> | <i>d</i> | <i>c+d</i> |
| | | <i>a+c</i> | <i>b+d</i> | <i>a+b+c+d</i> |
| | | 100 | 100 | 200 |

Figura 5. Relación entre un historial de inflamación de tobillos

| Características: | |
|--|---|
| $\frac{a}{(a+c)} = \frac{93}{100} = 0,93 = 93\%$ | Sensibilidad |
| $\frac{d}{(b+d)} = \frac{66}{100} = 0,66 = 66\%$ | Especificidad |
| Probabilidad pretest | |
| $\frac{(a+c)}{(a+b+c+d)} = \frac{100}{200} = 50\%$ | Prevalencia de la probabilidad pretest de tener el desorden estudiado (ascitis). |
| Predicciones: | |
| $\frac{a}{(a+b)} = \frac{93}{127} = 0,73 = 73\%$ | Valor predictivo positivo= Probabilidad postest de presentar el desorden estudiado (ascitis) en pacientes con historial de inflamación de tobillos. |
| $\frac{d}{(c+d)} = \frac{66}{73} = 0,90 = 90\%$ | Valor predictivo negativo =Probabilidad postest de no presentar el desorden estudiado (ascitis) en pacientes con historial negativo de inflamación de tobillos. |
| $\frac{c}{(c+d)} = \frac{7}{73} = 0,10 = 10\%$ | =Probabilidad postest de presentar el desorden estudiado (ascitis) en pacientes con historial negativo de inflamación de tobillos. |



El intervalo de confianza

Sin embargo, dado que este estudio solo incluyó 15 pacientes con ascitis (figura 4) bien se puede preguntar cuánto se puede confiar en la sensibilidad obtenida (0,93). Pues bien, el grado de confianza de este (o de cualquier otro) estimado de sensibilidad (o especificidad) puede ser calculado y expresado como un **"intervalo de confianza"** dentro del cual se puede confiar que se encuentra la verdadera sensibilidad, es decir, en el 95 % de las veces. En este caso, el intervalo de confianza del 95 % de esta sensibilidad de 0,93 basada en 15 pacientes recorre desde 0,81 (no muy

sensible) hasta 1,00 (o perfecta sensibilidad). Si, de otra parte, esta sensibilidad de 0,93 se basara en 100 pacientes con ascitis, el intervalo de confianza del 95 % iría desde 0,88 hasta 0,98 de forma que se estaría más confiado en que un historial negativo de inflamación de tobillos descarta la ascitis. Por lo anterior, siempre es importante buscar información sobre el intervalo de confianza del 95 % para medidas de exactitud tales como la sensibilidad y la especificidad.

UNA ESTRATEGIA MÁS PODEROSA Y MÁS RÁPIDA: LA RAZÓN DE VEROSIMILITUDES

| | | Presence of Ascites on Abdominal Ultrasound | | |
|---------------------------|-----|---|----------------------------|----------------------|
| | | Present | Absent | |
| History of Ankle Swelling | Yes | 14 (0.93) | 16 (0.33) | 30 <i>a+b</i> |
| | No | 1 (0.07) | 32 (0.67) | 33 <i>c+d</i> |
| | | <i>a+c</i> 15 (1.00) | <i>b+d</i> 48 (1.00) | <i>a+b+c+d</i> 63 |

Figura 6. Razón de verosimilitudes para un historial de inflamación de tobillos en el diagnóstico de ascitis.

El método de la razón de verosimilitudes usado para describir la exactitud de la información, es mucho más rápido y poderoso que la estrategia de sensibilidad y especificidad. La figura 6 muestra las razones de verosimilitud para un historial de inflamación de tobillos en el diagnóstico de ascitis. Resumiendo, la razón de verosimilitudes expresa la probabilidad de que un paciente con un hallazgo dado, en el historial y en el examen físico, tenga la enfermedad en oposición a otro paciente que no presente dicho hallazgo.

| Características: | |
|--|---|
| $\frac{\text{Sensibilidad}}{(1 - \text{especificidad})} = \frac{\left[\frac{a}{(a+c)}\right]}{\left[\frac{b}{(b+d)}\right]} = \frac{0,93}{0,33} = 2,8$ | Razón de verosimilitudes (de tener el desorden estudiado) para un resultado positivo (historial de inflamación de tobillos). |
| $\frac{1 - \text{sensibilidad}}{\text{Especificidad}} = \frac{\left[\frac{c}{(a+c)}\right]}{\left[\frac{d}{(b+d)}\right]} = \frac{0,07}{0,67} = 0,10$ | Razón de verosimilitudes (de tener el desorden estudiado) para un resultado negativo (historial negativo de inflamación de tobillos). |
| Probabilidad pretest | |
| $\frac{(a+c)}{(a+b+c+d)} = \frac{15}{63} = 24\%$ | Prevalencia de la probabilidad pretest de tener el desorden estudiado (ascitis). |
| Predicciones: | |
| <i>Historial positivo:</i> $\frac{0,24}{0,76} = 0,32 \times 2,8 = \frac{0,90}{1,90} = 47\%$ | Probabilidad posttest de presentar el desorden estudiado (expresado como razón de posibilidades) X razón de verosimilitudes para el resultado del test. |
| <i>Historial negativo:</i> $\frac{0,24}{0,76} = 0,32 \times 0,10 = \frac{0,03}{1,03} = 3\%$ | |

Cuando el cociente de probabilidad de un hallazgo está por encima de 1,0, la probabilidad de la enfermedad aumenta (debido a que los hallazgos se hacen más probables en aquellos pacientes que tienen el desorden).

Cuando el cociente de probabilidad está por debajo de 1,0, la probabilidad de presentar la enfermedad descende.

Finalmente, cuando el cociente de probabilidad está cerca de 1,0, la probabilidad de tener la enfermedad no varía (porque el hallazgo es igualmente probable tanto en pacientes con la enfermedad como en aquellos que no la tengan).



Ventajas de las razones de verosimilitud (Likelihood ratio)

Las razones de verosimilitud están relacionadas con la sensibilidad y la probabilidad, pero, para los clínicos, las primeras tienen algunas ventajas.

En una tabla de 2x2, tal como la de la figura 6, la **razón de verosimilitud para un historial positivo** de inflamación de tobillo es igual a:

$$\text{Sensibilidad}/(1-\text{especificidad}) \text{ o, } 0,93/0,33 \text{ o, } 2,8$$

Indicando así que un *historial positivo* es casi tres veces más probable en un paciente con ascitis que en un paciente sin este historial.

La **razón de verosimilitud para un historial negativo** de inflamación de tobillos es igual a:

$$(1-\text{sensibilidad})/\text{especificidad} \text{ o, } 0,07/0,67 \text{ o } 0,10$$

Indicando que un **historial negativo** solo es probable de ser encontrado 1/10 veces en un paciente con ascitis en oposición a un paciente sin ascitis.



La **primera ventaja** de las razones de verosimilitud es, cuando se aplica la probabilidad pretest de la enfermedad, genera la probabilidad posttest de tal enfermedad o desorden. Puesto que las razones de verosimilitud se expresan como una probabilidad, al principio puede parecer difícil de manejar, porque esto significa que la probabilidad pretest

debe también ser expresada como una probabilidad (más adelante se muestra cómo evitar hacer estos cálculos manualmente mediante el nomograma de la figura 7). Cuando se hace manualmente, la probabilidad (Odds) pretest de la enfermedad estudiada se convierte en una razón de probabilidad pretest mediante la siguiente fórmula:

$$\text{Probabilidad pretest} = \frac{\text{Probabilidad de tener la enfermedad (el desorden)}}{\text{Probabilidad de no tener la enfermedad}}$$

En la figura 6, la probabilidad de ascitis es de 0,24 y la probabilidad de no tener ascitis es 1,00-0,24 o, 0,76. Por lo tanto, la probabilidad (odds) pretest de ascitis es 0,24/0,76 o, 0,32 y esta puede ser multiplicada por 0,28 (generando una probabilidad posttest de 0,90) cuando es positivo el

historial de inflamación del tobillo y por 0,10 (generando una probabilidad posttest de 0,03) cuando el historial es negativo.

La anterior probabilidad posttest puede ser convertida a probabilidades con la fórmula:

$$\text{Probabilidad posttest} = \frac{\text{Probabilidad posttest}}{\text{Probabilidad posttest} + 1}$$

Así, la probabilidad posttest de 0,90 deducida de un historial positivo de inflamación de tobillo se convierte (al dividir 0,90/1,90) en 47 %, y probabilidad posttest de ascitis de 0,03 deducida de un historial negativo se convierte (al dividir 0,03/1,03) en 3 %, y estos son los mismos valores de la probabilidad posttest de tener ascitis generados en la figura 4.

Para un historial positivo de inflamación de tobillos: Colocar una regla en el margen izquierdo del nomograma sobre la probabilidad pretest de 24 % y rotarla hasta que intersecte la línea del centro del nomograma, donde se encuentran las razones de verosimilitud, en 2,8, correspondientes a un historial positivo de inflamación de tobillos; haciendo esto se intersectará también el margen derecho del nomograma justo debajo del 50 %.

La necesidad de convertir la probabilidad a razones y de nuevo invertir este proceso, se puede evitar usando el nomograma mostrado en la figura 7, el cual ya contiene las conversiones. Para hacerlo se debe:

Para un historial negativo de inflamación de tobillos: rotar la regla hasta que intersecte la línea de razones de verosimilitudes en 0,10, de esa forma la probabilidad posttest cae en 30 %.

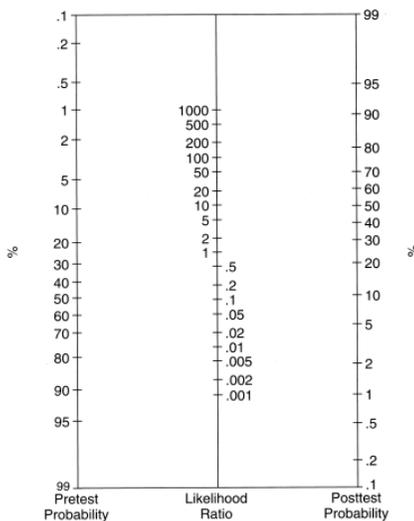


Fig 7.—A nomogram for applying likelihood ratios (adapted from reference 1).

La **segunda ventaja** de las razones de verosimilitud se vuelve evidente cuando se observa que el nomograma permite determinar la probabilidad de ascitis cuando la probabilidad pretest cambia de 24 % (figura 4) a 50 % (figura 5) sin tener que calcularla. Simplemente se vuelve a

colocar la regla en 50 % y se rota hasta que intersecte la línea de las razones de verosimilitud en 2,8, obteniendo de esa manera una probabilidad posttest de 73 %. Igualmente si se rota a 0,10, la probabilidad posttest cae en 10 %.

Table 2.—Multiple Levels of Responses to the CAGE Questions for Alcohol Abuse or Dependency*

| No. of Positive Answersto the 4 CAGE Questions | Alcohol Abuse or Depen- dency† | | Likelihood Ratios |
|--|--------------------------------|------------|-------------------|
| | Yes | No | |
| 4 | 23 (0.20) | 0 (0) | ∞ |
| 3 | 37 (0.32) | 1 (0.002) | 160 |
| 2 | 28 (0.24) | 14 (0.03) | 8 |
| 1 | 11 (0.09) | 28 (0.07) | 1.3 |
| 0 | 18 (0.15) | 358 (0.89) | 0.17 |
| ... | 117 (1.00) | 401 (1.00) | ... |

*Adapted from reference 5.

†Numbers in parentheses are proportions of the respective columns.

La **tercera ventaja** de las razones de verosimilitud es que, a diferencia de la sensibilidad y la especificidad (las cuales se limitan el número de resultados del test a dos niveles, “positivo” y “negativo”), pueden ser generados para múltiples niveles de resultados del test de diagnóstico. En cada nivel, la proporción de pacientes con la enfermedad es dividida entre la proporción de pacientes que no presentan la enfermedad en el mismo nivel; el resultado es el cociente de probabilidad para dicho nivel. Esto se muestra en la **tabla 2**, donde se muestran las razones de verosimilitud para 4, 3, 2, 1 y ninguna respuesta positivas al test de CAGE (para evitar la relación infinitamente alta del ítem de cuatro respuestas positivas se combinan los niveles de tres y cuatro respuestas positivas generando así una razón de verosimilitud de 255 para dicha combinación).

La **cuarta ventaja** de las razones de verosimilitud es que el posttest de probabilidad del desorden estudiado del primer dato de la información diagnóstica (por ejemplo, un historial de inflamación del tobillo) es el pretest de probabilidad de dicho diagnóstico para el próximo dato de la información diagnóstica (por ejemplo, el examen físico para el edema de tobillo).

Esta es una ventaja dado que los resultados del examen físico y de la información diagnóstica, incluyendo el laboratorio clínico y de radiología, no son independientes. Por ejemplo, un paciente con historial de inflamación de tobillos es más probable que tenga edema pedal que uno con historial negativo y este método permite emplear razones de verosimilitud para estos datos diagnósticos combinados en pareja o modificar la razón de verosimilitudes del segundo dato basados en los resultados del primero.



EN CONCLUSIÓN

Este artículo ha mostrado cómo se puede estimar la precisión y exactitud de los datos diagnósticos que se derivan de la historia clínica y del examen físico. Concluye con una tabla de trabajo (figura 8) y con un glosario que pueden ser de ayuda para evaluar las pruebas diagnósticas.

| | | Target Disorder | | |
|------------------------|----------|-----------------|--------|---------|
| | | Present | Absent | |
| Diagnostic Test Result | Positive | a | b | a+b |
| | Negative | c | d | c+d |
| | | a+c | b+d | a+b+c+d |

GLOSARIO

Para exactitud:

| | |
|--|--|
| $Sensibilidad = \frac{a}{(a+c)}$ | Cuando la sensibilidad es alta , un resultado de test negativo descarta la enfermedad estudiada. |
| $Especificidad = \frac{d}{(b+d)}$ | Cuando la especificidad es alta , un resultado de test positivo confirma la enfermedad estudiada. |
| $Valor\ predictivo\ positivo = \frac{a}{(a+b)}$ | Probabilidad postest de tener la enfermedad estudiada entre los pacientes con resultados de test positivos. |
| $Valor\ predictivo\ negativo = \frac{d}{(c+d)}$ | Probabilidad postest de no tener la enfermedad entre los pacientes con resultados negativos. |
| $\frac{c}{(c+d)}$ | Probabilidad postest de tener la enfermedad para pacientes con resultado de test negativo. |
| $Prevalencia = \frac{a+c}{(a+b+c+d)}$ | Probabilidad pretest de tener la enfermedad. |
| * $Cociente\ de\ probabilidad\ positiva = \frac{Sensibilidad}{1 - Especificidad}$ | Cociente de probabilidad (razón de verosimilitud positiva) de tener la enfermedad al tener un resultado de test positivo = $\frac{\frac{a}{a+c}}{\frac{b}{b+d}}$ |
| * $Cociente\ de\ probabilidad\ negativo = \frac{(1 - Sensibilidad)}{Especificidad}$ | Cociente de probabilidad (razón de verosimilitud negativa) de tener la enfermedad al tener un resultado de test negativo = $\frac{\frac{c}{a+c}}{\frac{d}{b+d}}$ |
| $Probabilidad\ postest\ de\ tener\ el\ trastorno\ objetivo\ (expresado\ como\ probabilidades)$ $=\ probabilidad\ pretest\ de\ tener\ el\ desorden\ (expresado\ como\ probabilidades)$ $\times\ Cociente\ de\ probabilidad\ para\ el\ resultado\ del\ test$ | |

Para precisión (y κ):

$$Concordancia\ observada: \frac{(a+d)}{(a+b+c+d)} = Concordancia\ esperada:$$

$$Celda\ a: \frac{([a+b] \times [a+c])}{(a+b+c+d)} =$$

$$Celda\ d: \frac{([c+d] \times [b+d])}{(a+b+c+d)} =$$

$$\frac{(esperado\ a + esperado\ b)}{(a+b+c+d)} =$$

$$Concordancia\ no\ debida\ al\ azar = \frac{\kappa (concordancia\ observada - concordancia\ esperada)}{(100\% - concordancia\ esperada)} =$$

Niveles convencionales para κ

| | | | | |
|-----------|-----------|-----------|---------------|---------------|
| Leve | Adecuado | Moderado | Significativo | Casi perfecto |
| 0,0 - 0,2 | 0,2 - 0,4 | 0,4 - 0,6 | 0,6 - 0,8 | 0,8 - 1,0 |

