

Contribución al estudio de pruebas diagnósticas: sobre sus índices fijos y valores predictivos

Patricio Herrera L.¹; Rosanna Lagos Z.²

Predictive values of diagnostic tests: potential distortions

Diagnostic tests are common and valuable tools in clinical practice, their efficiency being assessed by means of sensitivity, specificity and predictive values, derived after comparing the test results with a gold standard. Although basic guidelines to assess tests validity are well known, a published report on tests for detection of urinary tract infection in 990 children was hereby used as a model to review the implications of some requirements summarized in the guides for test assessment with the so called fixed indexes. After examination of the whole and of three clinical categories of disease fixed indexes, variation of the latter poses the question of whether the overall indexes or any of those of either category should be used for calculation of predictive values. Source of potential distortions or misuse of tests and of their predictive values are shown to deal with disease spectrum - as regarded from the disease natural history standpoint - as well as the strategies designed for patients referral at the time of test's assessment and the objectives of its further clinical usage.

(**Key words:** diagnostic tests, biometry, statistics.)

Las llamadas genéricamente "pruebas diagnósticas" son un instrumento común e importante en la práctica clínica. El término "eficiencia" de una prueba diagnóstica se refiere a su capacidad de indicar presencia o ausencia de enfermedad y los cuatro índices en que se expresa (sensibilidad, especificidad, valores predictivos positivo y negativo) suelen ser calculados de una tabla de 2×2^1 (tabla 1).

Un requisito esencial de la "sensibilidad" y la "especificidad" es que sus valores sean fijos (índices fijos) para una determinada prueba diagnóstica, de modo que la generalización de su uso en la solución de problemas que implican decisiones clínicas permita el ajuste apropiado de los valores predictivos de la prueba, en diversas situaciones epidemiológicas y clínicas que pueden resumirse en el término "prevalencia" o "probabilidad previa a la prueba", que

es la proporción real de la enfermedad o condición clínica estudiada en la población total a la que ésta es aplicada². La confirmación o descarte de la presencia de la condición o enfermedad se hace aplicando un patrón de referencia diagnóstica o *gold standard*. En otras palabras, para que los valores predictivos (índices variables) que originarán decisiones clínicas sean válidos, es necesario que los índices fijos (sensibilidad y especificidad) sean confiablemente estables o "fijos". Las bases metodológicas que permiten confiar en los valores de sensibilidad y especificidad de una prueba diagnóstica (tabla 2) han sido resumidos y comentados extensamente³.

Tempranamente en el uso de las pruebas diagnósticas, en el formato del análisis bayesiano, se advirtió la conveniencia de tener presente que los conceptos matemáticos asociados a las ideas de sensibilidad y especificidad no representan, necesariamente, un correlato estricto de las realidades biológica y de la medicina clínica⁴. No es raro que el desconocimiento de lo anterior y la desaprensión respecto a los requisitos para la correcta conducción del análisis de una prueba diagnóstica provoquen pro-

1. Universidad de Chile, Facultad de Medicina, Unidad de Epidemiología Clínica y Departamento de Pediatría (Campus Norte).

2. Servicio de Salud Metropolitano Norte. Hospital Roberto del Río. Apoyo financiero Fondecyt Proyecto 1940049.

Tabla 1

Interpretación de los resultados
de una prueba diagnóstica

Resultado de la prueba	Condición de la enfermedad		Total
	Presente	Ausente	
Positivo	a	b	a + b
Negativo	c	d	c + d
Total	a + c	b + d	a + b + c + d

a: verdaderos positivos; b: falsos positivos; c: falsos negativos; d: verdaderos negativos.

Índices fijos: sensibilidad: $a/(a+c)$ especificidad: $d/(b+d)$.

Índices variables: Valor predictivo positivo: $a/(a+b)$ Valor predictivo negativo: $d/(c+d)$

Prevalencia: $(a+c) / (a+b+c+d)$

blemas por disparidades entre autores con respecto a los valores de sensibilidad y especificidad de una misma prueba⁵. Con el fin de documentar e ilustrar el significado y los problemas que plantean algunos de los requisitos esquematizados por Sackett *et al.*³ y los efectos potenciales de ignorarlos, cumplirlos parcialmente o sobresimplificarlos, utilizaremos un artículo⁶ que significó un extenso debate sobre la materia entre sus autores.

El problema suscitado es una disyuntiva práctica: cuando se mide el comportamiento de una prueba diagnóstica en una población determinada, compuesta —por razones de espectro clínico— de subgrupos sindrómicos distin-

tamente diferentes, donde los índices fijos del análisis del conjunto difieren de aquéllos de los subgrupos, ¿cuáles índices fijos deben emplearse en el cálculo de los valores predictivos para un niño cualquiera del universo total de los pacientes? ¿Deben calcularse en base a los valores correspondientes a la clasificación del paciente en particular o del grupo sindrómico en que pueda ser ubicado? Si ha de procederse en esta segunda forma, ¿que asegura que el grupo total está constituido sólo por grupos discretos —no sobrepuestos uno al otro— y no por un número mayor, indeterminable en la práctica, resultante de las posibles combinaciones de los tres síndromes y otros posibles originados por la misma enfermedad?

Método

En el artículo en referencia se analizó el comportamiento de tres pruebas diagnósticas para la detección de infección urinaria (ITU) en 990 niños, como parte de un estudio más amplio sobre esta enfermedad. Con el fin de cumplir el segundo requisito de la guía de Sackett (tabla 2) se definieron, para la pesquisa de la ITU, tres grupos sindrómicos de niños elegibles para incorporar pacientes al estudio: grupo A, con signos clínicos de infección urinaria; grupo B, sólo con fiebre, y grupo C, con lo que se denominó signos indirectos de ITU, como enuresis secundaria, mal incremento ponderal. Además, en previsión de análisis en subgrupos (que, generalmente, no se incluyen en este tipo de estudios), se procuró que no hubiese sobrepuestos entre ellos. La base de datos construida para el estudio de infección urinaria permitió determinar los diversos índices de las pruebas aplicadas al conjunto de los

Tabla 2

Guía para decidir sobre la validez de una nueva prueba diagnóstica*

¿Hubo una comparación independiente, ciega, con un patrón diagnóstico de referencia?

¿Incluyó la muestra de pacientes un espectro adecuado de los enfermos y de los individuos con enfermedades diferentes pero que se pueden confundir con los primeros?

¿Se describió adecuadamente el lugar y filtro de referencia de los pacientes en que se hizo el estudio?

¿Se determinó la precisión y reproductibilidad de la prueba?

¿Se definió aceptablemente el término "normal"?

Si la prueba se propuso como parte de un conjunto o secuencia, ¿se determinó su contribución a la validez global del conjunto o secuencia?

¿Se describió la táctica para realizar la prueba en detalle suficiente para poder reproducirla exactamente?

¿Se determinó la utilidad o aplicabilidad real de la prueba?

* Adaptado de Sackett DL., Haynes RB., Tugwell P. *Clinical Epidemiology*. Little Brown and Co. Boston, 1985: 293.

sujetos y algunos análisis ulteriores. Para los fines de esta comunicación se compararon los índices variables resultantes del cálculo a partir de los índices fijos del grupo total, con los índices variables obtenidos a partir de los correspondientes índices fijos de cada uno de los subgrupos.

La sensibilidad y especificidad globales y calculadas para cada uno de los grupos por separado aparecen en la tabla 3. Los valores predictivos calculados a partir de los índices fijos aplicados a los subgrupos según categoría sindromática (estimación 1) y al grupo completo (estimación 2) aparecen en la tabla 4, donde puede observarse que existen discrepancias entre los calculados del grupo total y los de los subgrupos.

Comentario

Por razones de espacio, y con el fin de presentar en forma clara el examen del problema y su análisis, lo desglosaremos en varios rubros sin que, necesariamente, su orden implique una jerarquización por grado de importancia.

En primer término cabe señalar que el conocimiento de la historia natural de la enfermedad es indispensable para definir cual es el "espectro adecuado" (segundo requisito de Sackett) de ella. Generalmente, lo que consideramos "enfermedad" es un conjunto de signos o síntomas detectables, por el paciente o el médico, pero fácilmente escapan al intento de definir un "espectro amplio" las etapas de la entidad en que tales indicios no existen, p.e.: formas asintomáticas o subclínicas (que son causa del sesgo de pesquisa). En la muestra de 990 niños del estudio, una variedad de síndromes sugirieron al médico del nivel primario la existencia de ITU,

los que se ordenaron en tres subgrupos, desde el síndrome más obvio hasta el más sutil, pero de algún modo detectable. Cuando se trata de un tamizaje (*screening*), la muestra en estudio debe incluir —por este motivo— individuos aparentemente sanos, lo que no corresponde al modelo que examinamos, ya que los niños reclutados para este estudio estaban distintamente más o menos enfermos. En el tamizaje, los "casos" son los individuos con una cierta enfermedad o alteración. Los "controles" son los individuos sin ella. En la población que analizamos, los "casos" eran los niños con ITU, mientras los "controles" eran aquellos que la sugirieron al médico que los derivó, sin que el patrón de referencia diagnóstica pudiese comprobarla. Por esta razón, las pruebas en cuestión no son de tamizaje sino de "preselección" para confirmación diagnóstica. Si las aplicamos a una población sin requisitos de derivación (tamizaje en población general), en contraposición con los casos de nuestra muestra, los valores de los índices fijos (y, por ende, los valores predictivos) tendrán que diferir de todos los calculados en este estudio, lo cual sugiere que el objetivo de las pruebas diagnósticas determina, entre otras cosas, el tipo de universo que debe considerarse para el estudio de su eficiencia.

Al examinar las variaciones de los índices fijos resultantes de agrupar los niños del estudio, puede deducirse un efecto de este tipo si se considera que está constituido por grupos en los que la probabilidad de que el diagnóstico clínico coincida con el del patrón de referencia

Tabla 3

Sensibilidad y especificidad (%) de tres pruebas diagnósticas en 990 niños con infección del tracto urinario: efectos de subagrupar los casos según causa de sospecha*

Prueba	Grupo							
	Total (n: 990)		A (n: 569)		B (n: 252)		C (n: 169)	
	Sensib.	Especificidad	Sensib.	Especificidad	Sensib.	Especificidad	Sensib.	Especificidad
EL	83,0	83,3	85,3	85,5	79,4	79,8	42,8	84,5
N	45,0	95,8	46,6	92,2	44,1	99,1	15,4	97,4
ONL	91,1	71,8	93,0	77,3	85,3	63,3	54,5	74,2

*: adaptado de referencia 6.

Sensib.: sensibilidad; EL: esterasa leucocitaria; N: nitritos; ONL: orina no lúmpida.

Grupo A: niños con signos clínicos de infección urinaria; Grupo B: niños sólo con fiebre; Grupo C: niños con signos indirectos de infección urinaria.

Tabla 4

Valores predictivos de tres pruebas para la detección de infección del tracto urinario (ITU) en niños, en subgrupos clasificados según motivos de sospecha

Grupo:	Valores predictivos (%)			Negativo	
	Positivo		Prevalencia	Estimación 1*	Estimación 2**
Prevalencia	Estimación 1*	Estimación 2**		Estimación 1*	Estimación 2**
Esterasa leucocitaria					
A	86,8	84,7	83,9	81,5	52,7
B	38,0	43,7	96,1	96,8	13,5
C	31,6	31,6	94,2	98,5	8,3
Sensibilidad : 83,0%					
Especificidad: 83,3%					
Nitritos					
A	86,9	92,5	60,8	61,1	52,7
B	88,2	62,5	91,9	91,7	13,5
C	42,8	46,2	93,2	94,9	8,3
Sensibilidad : 45,4%					
Especificidad: 95,8%					
Aspecto no límpido					
A	82,0	78,2	90,8	87,7	52,7
B	26,6	33,7	96,5	98,1	13,5
C	18,3	22,8	95,8	99,1	8,3
Sensibilidad : 91,1%					
Especificidad: 71,8%					

A: Grupo con signos de ITU; B: grupo sólo con fiebre; C: grupo con signos indirectos de ITU.

*: los valores están calculados a partir de la sensibilidad y especificidad de cada grupo, después de la clasificación (no incluidos).

**: los valores están calculados a partir de la sensibilidad y especificidad en el grupo de 990 niños (al pie de cada tabla).

diagnóstica es claramente diferente. Hay, pues, un factor cercano aunque no idéntico al que significa la prevalencia (que, de hecho, puede suponerse diferente en cada uno de los grupos). Una primera conclusión (en cierto modo ingenua, porque de alguna manera hay que definir la población en riesgo, por ejemplo, de ITU) podría ser que el espectro por definir no debe fragmentarse en categorías. Esta conclusión se ve apoyada por el hecho predecible que la fragmentación del universo que constituye el "espectro adecuado" de la enfermedad implica una reducción, proporcional al número de categorías sindrómicas, lo que reducirá la precisión de los índices hasta límites cercanos a su inutilidad, por razones matemáticas.

Por otro lado, si el criterio de inclusión de los niños para este estudio de las pruebas se

hubiese restringido a sólo uno de estos grupos (p.e.: previsiblemente, los con ITU más evidente), los valores predictivos de estas pruebas se habrían distorsionado en forma considerable al generalizar los índices fijos del grupo en particular al espectro completo de la enfermedad, representado mejor por el total de pacientes, por favorecerse el reconocimiento de una estimación "promedio" (ver más adelante).

Como la práctica exige una definición de algún tipo, puede proponerse que, para no utilizar la(s) prueba(s), la decisión de basar el cálculo de los valores predictivos en los índices fijos del grupo total o los de cada subgrupo, para decidir en un caso en particular, depende del objetivo con que se utilizará(n) la(s) prueba(s), como también las consecuencias, gravedad o precio de las falsas predicciones, positiva y ne-

gativa. El estudio que nos sirve de modelo ilustra, por medio del análisis de categorías sindromáticas y los correspondientes índices fijos, en qué aspecto la sobresimplificación del segundo requisito de Sackett (tabla 2) puede inducir a error: no prever que "la amplitud del espectro" puede involucrar heterogeneidad de condiciones biológicas.

Otra forma de incurrir en este error es restringir el criterio de "casos" sólo a individuos con un grado de expresión tan extremo de la enfermedad, que el patrón de referencia diagnóstica no es necesario, hasta el punto de permitir una decisión —en este caso terapéutica— a ciegas (p.e.: niños con todos los signos y síntomas del síndrome inequívoco de ITU o, en el caso de pruebas etiológicas para enfermedades infecciosas, el uso de cultivos puros del agente por detectar, más que muestras de pacientes); a su vez, utilizar como "controles" aquellos pacientes en los que no cabe duda sobre la ausencia de la enfermedad o condición en estudio (p.e.: si el problema es la ITU, niños con sarampión, o en el caso de los estudios etiológicos, líquidos estériles). Este último tipo de error no es raro, sobre todo cuando no existe un patrón de referencia diagnóstica o *gold standard* seguro. En este caso, se confunde el sentido del término "sensibilidad" con el que tiene en otros aspectos del trabajo científico, particularmente el de un laboratorio de química: "capacidad de un método analítico para detectar cantidades muy pequeñas de cierta sustancia". Del mismo modo, "especificidad" puede también significar "la capacidad de un método analítico de detectar una sola sustancia y no otras"³. El primer corolario, obviamente no original, es que el conjunto de sujetos en los que se va a probar una prueba diagnóstica debe, necesariamente, incluir el espectro más amplio posible no sólo de los casos con la enfermedad en estudio, sino también de la gama de casos sin la enfermedad pero, de alguna manera, confundibles con ella¹.

En lo que respecta al tercer requisito (lugar y filtro de referencia) de Sackett, este aspecto tiene estrecha relación con lo anotado en el párrafo anterior. En el estudio utilizado como modelo, la categorización buscaba, precisamente, evitar el sesgo "filtro de referencia", más probable en un centro especializado en pacientes con nefropatías (y tendencia natural a concentrar cierto tipo de pielonefritis) que en un consultorio

abierto con el objetivo de dicha investigación. Aun sin ser este un problema de este estudio, la demostración de valores fijos inestables debidos a categorías de la enfermedad permite señalar que, en la fase de recolección de casos para estudiar una prueba diagnóstica, es indispensable prever que la categorización mal controlada, inducida o espontánea, de pacientes puede ser determinante de errores en las predicciones, por los factores analizados.

En lo referente al cálculo de los valores variables según la prevalencia, al momento de aplicar cualquiera de las tres pruebas analizadas a una población pediátrica, lo que correspondería, según el método ortodoxo, sería: conocer la prevalencia de ITU en esa población y, luego, calcular los valores predictivos en base a la sensibilidad y especificidad de esa prueba, de acuerdo a un objetivo previamente definido. Es este objetivo el que indicará si los índices fijos a usar serán los del grupo total o los de alguno de los subgrupos.

Las prevalencias en los subgrupos aparecen como diferentes entre sí, lo cual puede ser una realidad objetiva o un simple artificio; sin embargo, la primera premisa es válida sólo si se acepta que la definición de tres grupos discretos (en este ejemplo) agota el espectro de la enfermedad (lo que se cumplió razonablemente en el estudio utilizado como modelo). En otras palabras, si se supone que no existen pacientes que compartan rasgos de dos o más de los síndromes, o bien, que no hay más categorías de la enfermedad que las estipuladas entre las dos extremas. El análisis de la historia natural de la enfermedad sugerirá que tal suposición es, muchas veces, falsa. En tal caso, es difícil aceptar una (o tres) prueba(s) diagnóstica(s) destinada(s) a reconocer, de una población no sana, cuya sensibilidad y especificidad varíe dependiendo de la clasificación previa de cada paciente, ya seleccionado —sólo en base a criterios clínicos— como individuo en riesgo de la enfermedad específica. El examen de los subgrupos parece reconocer sólo variaciones de la prevalencia de la enfermedad. Es posible que el efecto de la categorización traduzca más que simples variaciones de ésta. Más bien cabe plantear que las diferencias observadas dependen de la probabilidad con que, en cada grupo, el síndrome afecte la eficiencia de las pruebas o del patrón de referencia diagnóstica, más que su sensi-

bilidad y especificidad, por tratarse, como grupo total, de una suma de poblaciones con diferencias respecto a la naturaleza de la enfermedad, tanto en los "casos" como en los "controles". En esta instancia no se trataría de una simple diferencia de prevalencia, sino de insuficiencia del patrón de referencia diagnóstica como tal (este aspecto se discute más adelante). La tendencia de los varones lactantes menores a originar muestras de orina con cultivos positivos a *Proteus sp.*, sin que haya ITU, es conocida e ilustra este aspecto del problema. En estas circunstancias no sería válido generalizar índices fijos y ajustar los valores predictivos en base a las diferentes "prevalencias" observadas en los subgrupos.

Frente a las interrogantes expuestas se pueden proponer otras consideraciones, una de las cuales es entender los índices (particularmente los índices fijos) como meras estimaciones de los valores reales (teniendo a la vista el significado de "reales", desde la perspectiva de la noción de historia natural de la enfermedad y las limitaciones propias de las definiciones clínicas), lo cual es coherente con el concepto que afecta a la mayoría de los supuestos en que se basa el conocimiento científico y, muy en particular, el basado en la probabilística, como el trabajo en medicina. Contemplar los índices fijos como estimaciones más o menos sesgadas de los valores propios o intrínsecos de la prueba permite entender que los requisitos de validez enumerados en la tabla 3 tienen por objeto contribuir a reducir el sesgo de tal estimación para hacerla lo más confiable posible, a la vez que operacional.

Desde un punto de vista cualitativo, si se observan los valores del artículo en referencia, puede advertirse que, aunque los de sensibilidad y especificidad parecen seguir una gradiente desde el grupo más obviamente asociado a pielonefritis (grupo A) hasta el más alejado (grupo C), el resultado de calcular estos índices a partir del grupo total (990 pacientes) no es un simple "promedio" de los índices de los tres grupos, pero depende de ellos. De aquí puede deducirse que, mientras mejor representada esté la enfermedad en sus etapas a través de la población elegida, más exactos serán los índices fijos. En principio, la simple extrapolación de sensibilidades y especificidades propuestas para pruebas diagnósticas deben merecer ciertas reservas si se desconoce la composición de la po-

blación en que fue analizada tal prueba. Esto aconsejaría un esfuerzo por verificar la prueba en la población local, en condiciones que aseguren su uso correcto. En el orden cuantitativo, se puede intuir que el sesgo en la estimación de los índices fijos tenderá a aumentar en razón inversa al tamaño de la muestra hasta el punto de invalidar tales estimaciones, tanto por razones de la estructura del grupo en aspectos biológicos y clínicos, como por razones meramente matemáticas.

Cabe comentar, aún, dos aspectos importantes a tener en cuenta y que merecen una discusión más extensa que lo que permite el objetivo de este análisis. El primero dice relación con el patrón de referencia diagnóstica. Rara vez se cuenta con un patrón de referencia diagnóstica perfecto, lo cual condiciona frecuentemente distorsiones en los índices de las pruebas diagnósticas.

El que la probabilidad de detectar la existencia real de enfermedad sea diferente en diversos segmentos del espectro clínico de ella no justifica, de por sí, las diferencias de sensibilidad y especificidad en los grupos clínicos que representan tales segmentos. Es atributo de una buena prueba tener índices fijos independientes de la prevalencia, hasta el punto de ser el eje para el cálculo determinado, precisamente, por diferencias de ella en poblaciones diferentes, a condición que el patrón de referencia diagnóstica sea confiable. Las diferentes características clínicas de los segmentos del espectro pueden conllevar factores de error para el desempeño de la prueba o la eficiencia del patrón de referencia diagnóstica y ello expresarse como "inestabilidad" de los índices fijos de la prueba. Los defectos de la eficiencia de este patrón pueden ser dependientes o independientes de las deficiencias de la prueba en estudio y el sentido del sesgo introducido por este factor merece estudios especiales para mejorar la estimación de dichos índices⁷. Por ejemplo, en una categoría de ITU (pielonefritis aguda) la poliuria resultante de la reducción de la capacidad de concentrar puede influir en la eficiencia del patrón de referencia diagnóstica.

Aun cuando las pruebas diagnósticas pueden ser usadas para reconocer un grupo de individuos aparentemente sanos en riesgo de cierta enfermedad, detectar una cierta enfermedad en un grupo de pacientes ya reconocidos como po-

blación en riesgo (como podría ser el caso de nuestro ejemplo), para la confirmación diagnóstica de una enfermedad e, incluso, para medir efectos de terapias y establecer pronósticos, conviene recordar que la mayor utilidad de una prueba diagnóstica se alcanza cuando la prevalencia del fenómeno que se desea detectar es de 40 a 60%^{3, 8}. La importancia de los falsos positivos y falsos negativos variará según el objetivo de la prueba, como también la gravedad de las decisiones a que dé lugar. Por ejemplo, un falso negativo pesará considerablemente cuando la enfermedad en estudio compromete la vida o su calidad en un paciente. Se puede ilustrar el significado de las diferencias en la sensibilidad de pruebas diagnósticas, tal vez de un modo más extremo que las variaciones de los índices variables observados en la tabla 4, examinando la sensibilidad de los hemocultivos en situaciones de septicemia clínica. Si una inadvertencia atribuye a tales cultivos sensibilidad del 100%, el resultado es valor predictivo negativo también igual a 100%, lo que implica que un resultado negativo de esta prueba excluye la probabilidad de enfermedad bacterémica (septicemia) y, por consiguiente, se justificaría abstenerse de usar antibióticos en tales circunstancias. A su vez, un falso positivo puede ser un precio bajo cuando el resultado de la prueba conduce a más estudio diagnóstico o una terapia sin efectos adversos y de bajo costo, si ello implica cubrir un grave riesgo imposible de cuantificar o descartar de otro modo. Dependiendo de la naturaleza del problema, la elección de los índices fijos para el cálculo de los índices variables se hará en forma más o menos conservadora. En términos muy simplificados, puede agregarse, como elemento de orientación estratégica, el concepto que la relación entre el "precio" de la suma de los falsos positivos más el de los falsos negativos, por un lado, y la suma de los "precios" de los verdaderos positivos más el de los verdaderos negativos, por otro, dará una idea del peso que tendrá el resultado de la prueba en las decisiones clínicas finales (p.e.: diagnóstica, terapéutica o pronóstica), criterio que puede considerarse análogo al usado para seleccionar el mejor "punto de corte" en una prueba que se basa en la medición de una variable continua³.

Como puede verse, la respuesta a las interrogantes planteadas no es simple y requiere una medida no siempre prevista de reflexión además

de claridad de objetivos y previsión de consecuencias. Sin perder de vista las ventajas operacionales de determinada prueba diagnóstica, es el reconocimiento de la importancia y sentido de sus índices fijos, la sensibilidad y la especificidad, propios del comportamiento de dicha prueba en una población determinada y, al momento de aplicarla a un paciente en particular (para conocer los valores predictivos o índices variables), lo que debe condicionar las características de su análisis de eficiencia, el momento de su uso clínico y el marco estratégico dentro del cual deben interpretarse los resultados, como etapa de un proceso racional diseñado para resolver las incertidumbres propias del trabajo clínico. El exceso de mecanización de estos conceptos y procedimientos tan frecuentemente usados en clínica puede inducir errores que alcanzan la esfera de la discusión ética.

Resumen

Las pruebas diagnósticas son una herramienta común y útil en la práctica clínica, cuya eficiencia se mide a través de su sensibilidad, especificidad y valores predictivos, los cuales se calculan después de comparar los resultados de la prueba con un patrón de referencia. Aunque las guías básicas para medir la validez de una prueba son bien conocidas, se usó como modelo una publicación sobre el comportamiento de tres pruebas diagnósticas en 990 niños con infección del tracto urinario, para examinar las implicaciones que tienen algunas de las exigencias de tales guías en relación con los llamados índices fijos. Después de estudiar los índices fijos del grupo total y de cada una de las tres categorías clínicas de la enfermedad, las variaciones de éstos plantean la interrogante sobre cuáles índices usar para el cálculo de los valores predictivos de las pruebas. Se muestra que la fuente de distorsiones o mal uso de los valores predictivos en la medición de eficiencia de pruebas diagnósticas dependen del espectro de la enfermedad, teniendo presente la historia natural de ésta, además de las estrategias usadas en la referencia de pacientes al momento del estudio de la prueba y de los objetivos definidos para su uso clínico ulterior.

(Palabras clave: pruebas de diagnóstico, biometría, estadísticas.)

Agradecimientos

A los Dres. Gastón Duffau y Carlos Y. Valenzuela, por sus valiosas sugerencias y revisión del manuscrito en todas sus fases.

Referencias

1. *Ransohoff D, Feinstein A: Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. NEJM 1978; 299: 926-930.*
2. *Riegelman R: Studying a study and testing a test. Little Brown and Co., Boston, 1981.*
3. *Sackett D, Haynes R, Tugwell P: Clinical Epidemiology. Little Brown and Co. Boston, 1985.*
4. *Feinstein AR: On the sensitivity, specificity and discrimination of diagnostic tests. Clin Pharmacol Therap 1975; 17: 104-116.*
5. *Harris JM: The hazards of bedside Bayes. JAMA 1981; 246: 2602-2605.*
6. *Lagos R, Carter J, Herrera P: Utilidad de una tira reactiva y del aspecto macroscópico de la orina para descartar la sospecha clínica de infección del tracto urinario en niños ambulatorios. Rev Chil Pediatr 1994; 65: 88-94.*
7. *Valenstein P: Evaluating diagnostic tests with imperfect standards. Am J Clin Pathol 1990; 93: 252-258.*
8. *Weinstein MC, Fineberg HV: Clinical Decision Analysis, W.B. Saunders Co. Philadelphia, 1980.*

Esta publicación está disponible en copias de microfilms de 16 y 35 mm y microfichas de 105 mm, las que pueden solicitarse a:

University Microfilms International
300 North Zeeb Road
Ann Arbor, Michigan 48106, USA.

This journal is also available in 16 mm microfilm, 35 mm microfilm and 105 mm microfilm copies through

*University Microfilms International,
300 North Zeeb Road,
Ann Arbor, Michigan 48106, USA.*