

Evaluación de tecnologías diagnósticas y tipos de muestreos

Luis Carlos Orozco, Diana M. Camargo

Resumen

La presente revisión y actualización tiene como objetivo mostrar al investigador de evaluación de tecnologías diagnósticas los diferentes tipos de muestreo que se pueden utilizar para disminuir costos y riesgos, y la manera de realizar los cálculos de las características de las pruebas, a saber: sensibilidad, especificidad, valor predictivo positivo y valor predictivo negativo de una manera sin sesgo en una muestra representativa de la población en que la prueba sería utilizada y se dan algunas recomendaciones para evitar errores demasiado frecuentes en la evaluación de pruebas diagnósticas como el muestreo seudoretrospectivo.

Summary

For the researcher evaluating sampling diagnostic technology, the present revision and up-dating of diagnostic and sampling techniques has as its objective the demonstration of how they can be used to reduce costs and risks and how to carry out test characteristics' calculations, to establish sensitivity, specificity, positive predictive value and negative predictive value, in an unbiased way in a representative sample of the population in which the test will be used and to give recommendations to prevent errors which frequently occur in test evaluation, such as in pseudo-retrospective sampling.

La historia de la evaluación de las tecnologías diagnósticas se remonta a la propuesta de Yerushalmy (1) para evaluar los fotofluorogramas en el diagnóstico de la tuberculosis. Su desarrollo ha sido relativamente rápido, pero, no es comparable con el de la evaluación de las intervenciones de prevención, terapéuticas o de rehabilitación que se basan en el ensayo clínico controlado (ECC) (2).

El objetivo de la presente revisión es mostrar los diferentes tipos de muestreo que se pueden utilizar cuando se evalúa una prueba comparada con un diagnóstico de referencia y la manera de calcular la sensibilidad, la especificidad y los valores predictivos de la prueba sin ningún sesgo.

Muestreo de corte transversal

El primer libro sobre la evaluación de tecnologías diagnósticas, *Beyond normality: the predictive value and efficiency of medical diagnoses*, por R.S. Galen y S.R. Gambino (3), publicado en 1975, fue la base de todo lo que se ha realizado en las dos últimas décadas. En 1981, Griner y colaboradores publicaron una revisión del tema en *Annals of Internal Medicine* (4).

En estos dos documentos, se presentan todas las fórmulas para calcular las probabilidades denominadas: sensibilidad (S), especificidad (E) y valores predictivos positivos y negativos (VPP, VPN) para un muestreo de corte transversal, aquel en donde a todos los sospechosos de la

¹.Escuela de enfermería, Universidad Industrial de santander, Bucaramanga, Santander.

².Escuela de fisioterapia, Universidad Industrial de Santander, Bucaramanga, Santander.

entidad en estudio, se les practica, en forma independiente, el diagnóstico y la prueba que se está evaluando. Es decir, en una muestra de N_0 individuos representativos de la población, se realizan N_0 diagnósticos y en forma independiente N_0 pruebas.

Un ejemplo sería: en 1.000 sintomáticos respiratorios sospechosos de tuberculosis, se practican 1.000 cultivos de la muestra de esputo (diagnóstico o prueba de referencia) y 1.000 baciloscopias, prueba de la que se desea establecer la S, la E, el VPP y el VPN.

Con este tipo de muestreo, si se utilizan las fórmulas usuales para los cálculos de las probabilidades, se obtienen valores sin sesgo.

		DIAGNOSTICO			
		Cultivo		-	
B a c i l o s A)	P R U E B A	+	-		
		25 (VP)	5 (FP)	30 (Q)	
	E	-	945 (VN)	970 (Q')	
		50 (P)	950 (P')	1.000 (N ₀)	

$$VP=25/1.000=VP/N_0$$

$$FP=5/1.000=FP/N_0$$

$$FN=25/1.000=FN/N_0$$

$$VN=945/1.000=VN/N_0$$

$$S=VP/P=0,025/0,05=0,5$$

$$E=VN/P'=0,945/0,95=0,995$$

$$VPP=VP/Q=0,025/0,03=0,833$$

$$VPN=VN/Q'=0,945/0,97=0,97$$

$$P=\text{Prevalencia}=(VP+FN)/N_0$$

$$P'=1-P \text{ (complemento de P)}$$

$$Q=\text{nivel de la prueba}=(VP+FP)/N_0$$

$$Q'=1-Q \text{ (complemento de Q)}$$

Con base en el ejemplo anterior, podemos afirmar que la baciloscopia tiene una sensibilidad de 50%, una especificidad de 99,5%, un valor predictivo positivo de 83,3% y un valor predictivo negativo de 97% y, además, que la prevalencia (P) es de 5% y el nivel de la prueba (Q) de 3%.

La mayoría de los conceptos aquí expresados son de uso común, solo aparecen dos conceptos nuevos: el de muestra representativa de la población (N_0) y el de nivel de la prueba (Q).

Estos dos conceptos son tomados del reciente libro de Helena Chmura Kraemer, *Evaluating*

medical tests. Objective and quantitative guidelines (5) que, aunque más dedicado a evaluar un nuevo concepto sobre la manera de evaluar la calidad de las pruebas, no pasa por alto los diferentes tipos de muestreo y las limitaciones que las fórmulas ya mencionadas tienen para su análisis.

El concepto de nivel de la prueba (Q) es muy importante cuando se reconoce que la S y la E son probabilidades condicionales, la probabilidad de que algo ocurra dado que otro fenómeno existe, que describen el comportamiento de la prueba en relación con el diagnóstico; éste, en una muestra representativa de la población (N_0), son P y P', las proporciones de diagnósticos positivos y negativos; de la misma manera, los VPP y VPN son probabilidades condicionales que describen el comportamiento del diagnóstico en relación con la prueba; ésta, en una muestra representativa de la población (N_0), son Q y Q', las proporciones de las pruebas positivas y negativas.

De hecho, S, E, VPP y VPN son fenómenos en espejo, no es sino cambiar en la tabla presentada el diagnóstico por la prueba y la posición de los signos (+ a la derecha de prueba y abajo en diagnóstico) y la S se convertiría en el VPN y la E en el VPP.

Otros tipos de muestreo: retrospectivo y prospectivo

Bajo ciertas condiciones, la posibilidad de aplicar a todos los individuos (N_0) representativos de la población una prueba demasiado costosa o riesgosa no sería recomendable. En esta situación, a todos los individuos de la muestra representativa de la población (N_0) se les realiza el diagnóstico, con lo cual obtenemos P y P' sin sesgo. De los individuos con diagnóstico positivo, se toma una muestra aleatoria (N_1) y de los individuos con diagnóstico negativo se toma una muestra aleatoria (N_2) y a estas dos muestras se les aplica la prueba que se va a evaluar.

Por definición, $(N_1)+(N_2)<(N_0)$, lo cual disminuye costos y riesgos.

Para obtener la S, basta con calcular la proporción de individuos con prueba positiva entre los

individuos con diagnóstico positivo (N_1). Para la E, se hace lo mismo con las pruebas negativas entre los individuos con diagnóstico negativo (N_2). Estos dos estimadores, S y E, no presentan sesgo y junto con P y P' son las únicas probabilidades que podemos calcular directamente de los datos obtenidos. Sin embargo, con la S, E y P podemos, calcular indirectamente los VPP y VPN debido a que:

$$VP = P \cdot S$$

$$FN = P \cdot (1-S)$$

$$FP = P' \cdot (1-E)$$

$$VN = P' \cdot E$$

Con estos datos, los otros cálculos son fáciles.

$$Q = (P \cdot S) + (P' \cdot (1-E))$$

$$VPP = P \cdot S / Q$$

$$VPN = P' \cdot E / Q'$$

A este tipo de muestreo lo llama Kraemer (5) retrospectivo, debido a la similitud que tiene con la metodología de casos y controles, pero hay una diferencia: el realizar el diagnóstico en la totalidad de los (N_0) lo hace diferente porque se puede calcular la prevalencia (P). Esta sería una situación ideal para los casos y controles.

A veces resulta que el diagnóstico es más costoso o riesgoso que la prueba y lo que se puede hacer es un muestreo prospectivo (recuerda las cohortes), que es una imagen en espejo del retrospectivo. Este muestreo es útil en la evaluación de pruebas diagnósticas y es el único posible con pruebas pronósticas donde el diagnóstico se realiza tiempo después de practicadas las pruebas.

En el muestreo prospectivo, al total de individuos en la muestra representativa de la población (N_0) se les realiza la prueba. Así podemos obtener Q y Q' de manera directa sin sesgos y, a una muestra aleatoria (N_1) de positivos y (N_2) de negativos en la prueba, se les realiza el diagnóstico.

De nuevo, (N_1) + (N_2) < (N_0) .

La proporción de diagnósticos positivos entre los positivos en la prueba (N_1) es el VPP y la proporción de diagnósticos negativos entre los negativos en la prueba (N_2) es el VPN.

Q, su complemento Q' y los VPP y VPN son los únicos datos sin sesgo que se pueden obtener directamente con este tipo de muestreo. La S y la E se deben calcular indirectamente, debido a que:

$$VP = Q \cdot VPP$$

$$FN = Q' \cdot (1-VPN)$$

$$FP = Q \cdot (1-VPP)$$

$$VN = Q' \cdot VPN$$

Es fácil demostrar, entonces, que:

$$P = (Q \cdot VPP) + (Q' \cdot (1-VPN))$$

$$S = Q \cdot VPP / P$$

$$E = Q' \cdot VPN / P'$$

Los tres tipos de muestreo ya mencionados, son los únicos que producen resultados válidos, si se utilizan los cálculos recomendados.

Muestreos inadecuados

Con demasiada frecuencia se realiza un muestreo que Kraemer (5) denomina pseudo-retrospectivo el cual no produce estimaciones válidas. Este tipo de muestreo es aquel en que se utiliza un número de individuos con el diagnóstico positivo y un grupo, mal denominado control, que con mucha frecuencia son las personas del laboratorio o los estudiantes de la facultad o cualquier otro grupo de personas con un riesgo muchas veces más bajo que el de los posibles sospechosos del diagnóstico y a quienes jamás se les aplicaría la prueba que se está evaluando.

Este tipo de muestreo ha sido descrito también como fases II o III por Nierenberg y Feinstein (6), quienes demuestran con el análisis histórico del desarrollo de una prueba diagnóstica que los resultados en estas fases no dan resultados válidos.

El artículo mencionado (6) es una buena referencia para aquellas personas que trabajan en la estandarización y validación de tecnologías diagnósticas, porque muestra con bastante claridad los pasos que deben seguirse dentro de la misma. La debilidad de esta publicación es que

en la fase V, la definitiva para establecer los valores de la S, E, VPP y VPN, solo plantea un tipo de muestreo que, aunque lo denomina de tipo cohorte, es en realidad uno de corte transversal.

En resumen, cuando se realice un estudio para establecer S, E, VPP y VPN se debe utilizar uno de los tres tipos de muestreo recomendados: corte transversal, retrospectivo o prospectivo aplicando en su análisis las fórmulas enunciadas. El muestreo seudoretrospectivo no se debe utilizar más allá de una fase inicial de estandarización y, si se calculan los valores de S, E, VPP y VPN, debe anotarse que éstas no son probabilidades válidas. Jamás se debe poner en uso rutinario una prueba diagnóstica que solo ha sido evaluada con este tipo de muestreo.

Referencias

1. **Yerushalmy J.** Statistical problems in assessing methods of medical diagnosis with special reference to X-ray techniques. *Public Health Reports* 1947;62:1431-49.
2. **Anónimo.** The value of diagnostic tests. *Lancet* 1979; April 14:809-10.
3. **Galen RS, Gambino SR.** Beyond normality: the predictive value and efficiency of medical diagnoses. New York: John Wiley & Sons; 1975.
4. **Griner PF, Maywski RJ, Mushlim AI, Greenland P.** Selection and interpretation of diagnostic tests and procedures. *Ann Intern Med* 1981;94:553-600.
5. **Kraemer HC.** Evaluating medical tests. Objective and quantitative guidelines. Newbury Park: Sage Publications; 1992.
6. **Nierenberg AA, Feinstein AR.** How to evaluate a diagnostic marker test. Lessons from the rise and fall of dexamethasone suppression test. *JAMA* 1988;259:1699-1702.