

Universidade Federal de Minas Gerais

Instituto de Ciências Exatas

Departamento de Estatística

Estudio de la Calidad de Pruebas Diagnósticas

Ilka Afonso Reis

RTE02-2020

Resumen

Este informe académico tiene el objetivo de introducir algunos conceptos relacionados a las pruebas diagnósticas (sensibilidad, especificidad, valores de predicción positiva y negativa), que están presentados con más detalles en las secciones 3 y 4 del reporte técnico “*Avaliação de Testes Diagnósticos*” (Reis y Reis, 2002). Además de eso, son presentados los conceptos y el uso de las razones de verosimilitud positiva y negativa.

Como citar: Reis, I. A. (2020) Estudio de la Calidad de Pruebas Diagnósticas. Relatório Técnico do Departamento de Estatística da Universidade de Minas Gerais – RTE02-2020. Disponible en: <http://www.est.ufmg.br/portal/producao/relatorios-tecnicos/serie-ensino>

Índice

1. Introducción
2. Como Evaluar una Prueba Diagnóstica
3. Las Medidas de Calidad de la Prueba
4. Las Medidas de Calidad del Diagnóstico
5. Las Razones de Verosimilitud
6. Conclusiones
7. Bibliografía

1. Introducción

Una prueba diagnóstica es un procedimiento realizado para detectar una condición o evento de interés, es decir, una enfermedad, una infección, una contaminación, un caso de fraude, el uso de sustancias prohibidas y otros eventos más positivos como un embarazo o el género de un bebé.

Pruebas diagnósticas están ampliamente presentes en las áreas de Salud y Comportamiento. No obstante, también son empleadas en las áreas bancaria (detección de fraudes, concesión de préstamos), de *marketing* (identificación de posibles compradores para un producto) y hasta mismo en las áreas de vigilancia digital (detección de ocurrencia de crímenes por medio de inspección de *tweets*, reconocimiento facial) y de seguridad pública (detección de la presencia de sustancias ilícitas en equipajes, carros y otros escondrijos).

Las pruebas diagnósticas más conocidas son las pruebas clínicas, aquellas que necesitan de una colecta de materiales como sangre u orina, por ejemplo. Sin embargo, hay también las pruebas basadas en cuestionarios, que son utilizadas para detectar condiciones psíquicas como depresión, ansiedad y trastornos de atención, entre otras.

2. Como Evaluar una Prueba Diagnóstica

Las pruebas diagnósticas pueden acertar o errar la verdadera condición del individuo testado. En un individuo que posee la condición de interés, si la prueba resulta positiva, tenemos un verdadero positivo (VP). Por otro lado, si la prueba resulta negativa para ese individuo, tenemos un falso negativo (FN). Del mismo modo, en un individuo que no posee la condición de interés, si la prueba resulta negativa, tenemos un verdadero negativo (VN), pero, si resulta positiva, tenemos un falso positivo (FP).

Una prueba diagnóstica que tiene altísimas proporciones de verdaderos positivos y negativos, es decir, no yerra casi nunca, es llamada “patrón de oro”. Infelizmente, las pruebas de ese tipo suelen ser muy caras, demoradas o exigen procedimientos invasivos o personas muy expertas. De ese modo, su amplio uso en la población no es viable.

Antes de ser utilizada, una prueba diagnóstica necesita de una evaluación de su calidad. Para hacer esa evaluación, son necesarios dos grupos de individuos: los que tienen la condición de interés (los “enfermos”) y los que no tienen la condición de interés (los “sanos”). Si bien que la condición de interés no sea siempre una enfermedad, los términos “enfermos” y “sanos” son ampliamente usados en la literatura sobre el tema, ya que esa es dominada por aplicaciones en el área biomédica. Así es que también serán utilizados en este informe.

Para saber a cual grupo pertenece cada uno de los individuos, una prueba del tipo patrón de oro es utilizada. Además, los individuos también son clasificados en positivos o negativos por la prueba diagnóstica en evaluación. El cruzamiento de los resultados de ambas las pruebas suele ser presentado en una tabla como la Tabla 1, que presenta dos datos obtenidos por Pinho y Matos (2002) en su estudio de evaluación de la calidad del examen Papanicolaou. Esa es una prueba diagnóstica ampliamente utilizada para detectar cáncer cervical uterino. Para detectar ese tipo de cáncer, el patrón de oro suele ser la biopsia cervical, que es un examen más invasivo, caro y demorado.

Tabla 1: Resultados de la prueba Papanicolaou para detección del cáncer cervical uterino

Resultado de la prueba Papanicolaou	Cáncer cervical uterino		Total
	Enferma	Sana	
Positivo	265 (VP)	47 (FP)	312
Negativo	11 (FN)	50 (VN)	61
Total	276	97	373

Fuente: Pinho y Matos (2002).

En la Tabla 1, están identificados los verdaderos y falsos positivos (VP y FP, respectivamente), así como los verdaderos y falsos negativos (VN y FN, respectivamente).

3. Las Medidas de Calidad de la Prueba

La calidad de una prueba diagnóstica es medida por medio de dos proporciones de acierto: la sensibilidad y la especificidad. La sensibilidad (s) de una prueba es la proporción de resultados verdaderos positivos entre los enfermos y es calculada usándose la siguiente expresión

$$s = \frac{VP}{VP + FN} \cdot$$

De manera similar, la especificidad (e) de una prueba es la proporción de resultados verdaderos negativos entre los sanos y es calculada como se sigue

$$e = \frac{VN}{VN + FP} \cdot$$

En el ejemplo de la Tabla 1, el valor de la sensibilidad es $s=265/(11+276)=96,0\%$ y la especificidad es $e=47/(47+50)=51,5\%$. Así es que el examen Papanicolaou acierta más diagnósticos entre los enfermos que entre los sanos. En ese caso, decimos que el examen Papanicolaou es una prueba más sensible que específica.

La sensibilidad y la especificidad son características intrínsecas de una prueba diagnóstica, es decir, no cambian, aunque la cantidad de enfermos en una población aumente o disminuya.

4. Las Medidas de Calidad del Diagnóstico

Dado que un resultado positivo puede ser verdadero o falso, ¿cuál es la probabilidad que un individuo con resultado positivo sea realmente un enfermo? Esa probabilidad se llama Valor de Predicción Positiva (VPP) y es definida como la proporción de verdaderos positivos entre todos los positivos. Del mismo modo, si un individuo tiene un resultado negativo en la prueba, ¿cuál es la probabilidad que ese individuo con resultado negativo sea realmente un sano? Esa probabilidad se llama Valor de Predicción Negativa (VPN) y es definida como la proporción de verdaderos negativos entre todos los negativos. El VPP y el VPN son medidas de la calidad del diagnóstico basado en el resultado de la prueba.

En el ejemplo de la Tabla 1, los valores de predicción positiva y negativa son calculados como $VPP=265/312=84,9\%$ y $VPN=50/61=82,0\%$, respectivamente. En la situación presentada en la Tabla 1, el examen Papanicolaou es un poco más confiable cuando su resultado es positivo.

Ese resultado ocurre principalmente porque la proporción de enfermos en la muestra de la Tabla 1 es alta ($276/373=74,0\%$). Esa proporción de enfermos es conocida como prevalencia (p). En general, en estudios de evaluación de la calidad de pruebas diagnósticas, la prevalencia de la condición en la muestra es mucho más grande que la prevalencia de la condición en la población que usará la prueba.

Al contrario de la sensibilidad y la especificidad, que son propias de la prueba y no varían con la prevalencia de la condición en la población, los valores de predicción positiva y negativa dependen del valor de esa prevalencia. El VPP y el VPN son calculados por medio de las siguientes expresiones

$$VPP = \frac{s \times p}{s \times p + (1 - e) \times (1 - p)} \quad \text{y} \quad VPN = \frac{e \times (1 - p)}{e \times (1 - p) + (1 - s) \times p} .$$

Examinando las expresiones arriba, es fácil notar que tanto el VPP como el VPN dependen de la prevalencia (p) de la condición de interés en la población donde la prueba será aplicada. Caso el examen Papanicolaou sea aplicado a mujeres de un grupo cuya prevalencia de cáncer cervical uterino es $p=5\%$, por ejemplo, los valores de predicción serán $VPP= 9,4\%$ y $VNP=99,6\%$. En ese caso, el examen Papanicolaou será una prueba más útil para descartar la presencia del cáncer, pues un resultado negativo tendrá mucho más valor predictivo que un resultado positivo (VPN es más grande que VPP). Para confirmar los resultados positivos, las mujeres con ese resultado son sometidas a una segunda prueba, que tiene un VPP más grande, como la biopsia cervical.

El VPP y el VPN también dependen de la sensibilidad y de la especificidad de la prueba. Pruebas muy sensibles tienen un VPN grande, cualquiera sea la prevalencia de la condición, y, por eso, son útiles para descartarla. Pruebas muy sensibles son usadas como herramientas de cribaje, pues descartan los individuos negativos con gran confianza mientras retienen los individuos positivos para una prueba de confirmación. Por otro lado, pruebas muy específicas tienen un VPP grande, lo que las hace útiles para confirmar la condición, cualquiera sea su prevalencia. Entonces, pruebas muy específicas suelen ser usadas como pruebas de confirmación.

5. Las Razones de Verosimilitud

Como se ha mencionado anteriormente, la sensibilidad y la especificidad de una prueba diagnóstica la caracterizan totalmente y no dependen de la prevalencia de la enfermedad en la población testada. Una manera de unir esas dos medidas es calcular la razón de verosimilitud positiva (LR^+) y la razón de verosimilitud negativa (LR^-) de una prueba diagnóstica. La sigla LR se debe al termino en inglés *Likelihood Ratios*.

Las razones de verosimilitud son calculadas como se sigue

$$LR^+ = \frac{s}{(1-e)} \quad \text{y} \quad LR^- = \frac{(1-s)}{e}$$

La LR^+ compara la verosimilitud de un resultado positivo en una persona enferma con la verosimilitud de un resultado positivo en una persona sana. Es decir, “qué tanto más probable es encontrar la prueba positiva en alguien enfermo que en alguien sano” (Salech et al.,2008). En el ejemplo de la Tabla 1, el valor de la LR^+ es calculado como $LR^+ = 0,960/(1-0,515) = 1,979$. Así, la verosimilitud de un resultado positivo en una persona sana es casi dos veces la verosimilitud de ese resultado en una persona sana, o sea, es más probable que una prueba resulte positiva en una persona enferma que en una persona sana.

Por su vez, la LR^- compara la verosimilitud de un resultado negativo en una persona enferma con la verosimilitud de un resultado negativo en una persona sana. O sea, “qué tanto más probable es encontrar la prueba negativa en alguien enfermo que en alguien sano” (Salech et al.,2008). Con los datos de la Tabla 1, es posible calcular que $LR^- = (1-0,960)/0,515 = 0,078$. Eso significa que la verosimilitud de un resultado negativo en una persona sana es cerca de 8% de la verosimilitud de ese resultado en una persona sana, es decir, es mucho menos probable que una prueba resulte negativa en una persona enferma que en una persona sana.

Como ambas las LRs son definidas con los enfermos en el numerador de la razón y los sanos en su denominador, es deseable que el valor de LR^+ sea muy grande y que el valor de LR^- sea muy pequeño. De hecho, la LR^+ y la LR^- suelen ser usadas en la evaluación de la utilidad clínica de la prueba, es decir, si una prueba es más útil para confirmar la presencia de la enfermedad (prueba de confirmación) o más útil para descartar la presencia de la enfermedad (prueba de cribaje). Algunos autores afirman que una prueba que tenga $LR^+ > 10$ es muy útil como prueba de confirmación mientras una prueba que tenga $LR^- < 0,1$ es muy útil como prueba de cribaje (Deeks y Altman, 2004). En el caso del examen Papanicolau, el estudio de Pinho y Matos (2002) ha confirmado la utilidad ampliamente conocida de ese examen para descartar la presencia de cáncer cervical uterino en cualquier grupo de mujeres, ya que su valor de LR^- es 0,078.

Los valores de predicción positiva y negativa están relacionados a las razones de verosimilitud por medio de expresiones que son muy bien detalladas en el trabajo de Ranganathan y Aggarwal (2018). Esos autores demuestran cómo obtener las probabilidades *post-test* (valores de predicción) a partir de la probabilidad *pre-test* (prevalencia) y de las razones de verosimilitud.

Por lo general, una prueba con alto valor para LR^+ (>10) tendrá un alto valor de predicción positiva (VPP) cualquiera sea la prevalencia de enfermedad en la población testada y, por lo tanto, será útil como prueba de confirmación. En cambio, una prueba con bajo valor para LR^- (<0.1) siempre tendrá un alto valor de predicción negativa (VPN) y será útil como prueba de cribaje.

6. Conclusiones

La calidad de los resultados de una prueba diagnóstica puede ser evaluada por medio de medidas sencillas, que son nada más que proporciones y razones. Esa evaluación es necesaria para caracterizar el desempeño de la prueba en los grupos enfermo y sano, es decir, determinar su sensibilidad y su especificidad, respectivamente. La unión de esas dos medidas en las razones de verosimilitud positiva y negativa suele ser usada para evaluar si una prueba es más útil para confirmar la presencia de la enfermedad o para descartarla sin la necesidad de conocer su prevalencia. Por último, es importante saber cómo interpretar el resultado de una prueba diagnóstica, lo que es posible hacer por medio del entendimiento de los conceptos de los valores de predicción positiva y negativa.

7. Bibliografía

Deeks, J.J. ; Altman, D.G. (2004) Diagnostic tests 4: Likelihood ratios. **BMJ** v. 329, p.168-169

Pinho, A.A.; Matos, M.C.F.I (2002) Validade da citologia cervicovaginal na detecção de lesões pré-neoplásicas e neoplásicas de colo de útero. **Jornal Brasileiro de Patologia e Medicina Laboratorial**, v. 38, n. 3, p. 225-231.

Ranganathan, P.; Aggarwal, R. (2018) Understanding the properties of diagnostic tests – Part 2: Likelihood ratios. **Perspectives in Clinical Research**, v. 9, n. 2, p. 99–102.

Reis, E. A.; Reis, I. A. (2002) **Avaliação de Testes Diagnósticos**. Relatório Técnico do Departamento de Estatística da Universidade Federal de Minas Gerais – Série Ensino
Disponível em: <http://www.est.ufmg.br/portal/arquivos/rts/rte0203.pdf>

Salech, F. *et al.* (2008) Estudios que evalúan un test diagnóstico: interpretando sus resultados. **Revista médica de Chile**, v. 136, n. 9, p. 1208–1208.