

## EL PROBLEMA DE LA VALIDEZ DE UNA PRUEBA DIAGNOSTICA PARA USO MASIVO COMO PROCEDIMIENTO ESTADISTICO DE CLASIFICACION

Vicente M. Astudillo<sup>1</sup> & Isabel N. Kantor<sup>2</sup>

### RESUMEN

Se evalúa la efectividad de una prueba diagnóstica indirecta de uso masivo en procedimientos discriminantes a través de las probabilidades condicionales haciendo aplicación del teorema de Bayes. Se presentan ejemplos de estimaciones de los errores de diagnóstico y de la verdadera tasa de prevalencia.

Para estudios en poblaciones donde la presencia de animales positivos puede ser considerada un evento raro, se presenta la relación que existe entre el tamaño de la muestra y la probabilidad de fallar en detectar positivos.

### INTRODUCCION

Siempre que se utiliza una prueba indirecta para clasificar individuos de una población, en categorías de positivos y negativos con respecto a una enfermedad, es necesario evaluar la validez de la prueba. Es decir, saber si la prueba "mide" lo que se propone "medir". Por ejemplo, ¿la prueba de tuberculina en bovinos evalúa realmente la presencia o ausencia de tuberculosis? La respuesta a esta pregunta no es del tipo sí/no, sino que es dada en una escala proporcional o porcentual que indica su grado de validez.

Dos parámetros son utilizados para estudiar específicamente la validez de una prueba indirecta de clasificación diagnóstica aplicada a poblaciones: la sensibilidad y la especificidad. Estas pruebas casi nunca identifican exactamente todos los verdaderos positivos y negativos. Por ejemplo, la prueba de tuberculina produce positivos como

un resultado a la exposición al *Mycobacterium*. Ya que se toma el ejemplo de la tuberculosis conviene aclarar que, como respuesta inmunitaria, la reacción positiva del ganado bovino a la tuberculina PPD bovina es manifestación de una exposición anterior del animal al *Mycobacterium bovis* y, por tanto, significa infección y no necesariamente enfermedad. Aunque por la patogenicidad de este germen y la susceptibilidad de los bovinos a él, la infección va seguida por lesiones propias de la enfermedad en un plazo no muy largo. En otras ocasiones puede haber reacciones tuberculínicas positivas por infecciones debidas a otras micobacterias diferentes del *M. bovis*, por la existencia de antígenos comunes, que no van seguidas necesariamente de enfermedad en el bovino. Prácticamente esas reacciones para-específicas son las que adquieren mayor importancia relativa en las etapas de prevalencia muy baja de tuberculosis en bovinos (falsos positivos). En cambio, el error de diagnóstico opuesto, falsos negativos, puede ocurrir en los bovinos en las primeras fases de la infección (período pre-alérgico), durante otras infecciones intercurrentes (virosis), o en animales en estado caquéctico.

Los resultados falsos, llamados errores de clasificación, señalan el nivel de falibilidad de la prueba indirecta y además producen tanto dificultades en la estimación de la prevalencia de la tuberculosis, como en la toma de decisiones para la aplicación de medidas sanitarias tendientes a la erradicación de la enfermedad.

### EVALUACION DE LA VALIDEZ DE LA PRUEBA INDIRECTA

El desempeño de la prueba indirecta (tuberculización en este ejemplo) se puede evaluar a través de la sensibilidad y especificidad. En la práctica esto se puede hacer mediante un estudio de los datos de los animales de una región o país

<sup>1</sup>Centro Panamericano de Fiebre Aftosa, (OPS/OMS), Caixa Postal 589, 20000 Rio de Janeiro-RJ, Brasil.

<sup>2</sup>Centro Panamericano de Zoonosis, (OPS/OMS), Casilla 3092 - Correo Central, 1000 Buenos Aires, Argentina.

dado (R), donde existe un programa de control de la tuberculosis bovina. Para ello es necesario contar con la siguiente información: i) resultados de pruebas de tuberculinización para un cierto período (I), y ii) resultados de las inspecciones macroscópicas post-mortem en mataderos, exámenes histopatológicos y/o de pruebas de aislamiento en cultivos en el laboratorio (D).

En el pasado, en la región R hubo animales positivos a la prueba de tuberculina que fueron sacrificados en mataderos sanitarios, se inspeccionaron, se obtuvieron muestras y se analizaron en el laboratorio para aislar el agente. Por otro lado, animales que siempre fueron negativos a la prueba de tuberculina, en la región R, al llegar a cierta edad son descartados y enviados a matadero. Tales datos pueden servir ya que hay registro individual en el matadero con identificación del origen. Estos datos retrospectivos de la población-objeto de la región R permiten confeccionar la distribución de frecuencias (5) presentada en el Cuadro 1 y en la Figura 1.

La validez de una prueba diagnóstica debe reflejar su capacidad de proporcionar resultados verdaderos de aquello que se está evaluando. Para medir la validez de una prueba es necesario estimar su sensibilidad y su especificidad.

En el ejemplo de tuberculosis bovina que se ha dado, la sensibilidad (U) sería la capacidad de la prueba tuberculínica de indentificar como positivo a todo animal del rebaño que verdaderamente está infectado por *M. bovis*. Igualmente, la especificidad (V) sería la capacidad de la prueba tuberculínica

de identificar como negativo a todo bovino del rebaño que realmente no está infectado por *M. bovis*. Por las características de la especificidad, ella puede ser evaluada en áreas con rebaños libres de tuberculosis bovina.

CUADRO 1. Clasificación cruzada de las pruebas directas e indirectas. Denominación de las frecuencias celulares y marginales

Situación prueba indirecta	Situación real (directa)		
	D+	D-	Total
I+	a	b	(a+b)
I-	c	d	(c+d)
Total	(a+c)	(b+d)	(n)

donde:

- a = Número de animales considerados positivos por la prueba tuberculínica y que se hallan infectados, son los verdaderos positivos.
- b = Número de animales igualmente encontrados positivos a la prueba tuberculínica, en los que no se puede demostrar la infección por inspección post-mortem y cultivo de órganos. Se los denomina falsos positivos.
- c = Número de animales considerados negativos por la prueba tuberculínica, y que se hallan infectados. Se les denomina falsos negativos.
- d = Número de animales considerados negativos por la prueba tuberculínica, en los que no se puede demostrar infección, son los verdaderos negativos.
- a+c = Número total de animales infectados.
- b+d = Número total de animales no infectados.
- n = Número total de animales.

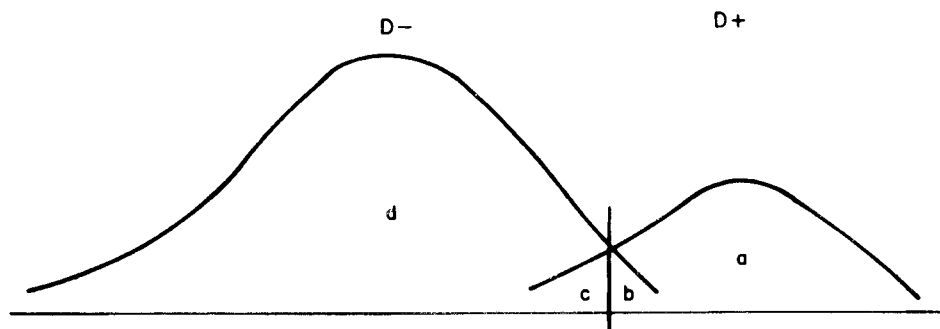


FIGURA 1. Distribuciones de frecuencias superpuestas de positivos y negativos. Zona de incertidumbre: errores de clasificación a través de prueba indirecta.

A partir de los datos del Cuadro 1 es posible estimar estos valores característicos:

$$U = a/(a+c)$$

$$V = d/(b+d)$$

La prueba tuberculínica simple en la tabla del cuello, utilizando tuberculina PPD bovina de gran potencia, llega a tener una sensibilidad óptima de 91% y una especificidad de 89% (3).

Al utilizar estos indicadores de validez es conveniente tener en cuenta que en regiones donde el porcentaje de infección por *M. bovis* es relativamente alto, puede haber animales con infección reciente, en que la prueba de tuberculina es positiva. Sin embargo, las pruebas directas pueden fallar en la confirmación, puesto que existe la posibilidad de que aún no se haya presentado lesión macroscópica. Por tal razón, la muestra de ganglios y órganos para cultivo puede no llegar a incluir tejido donde la primoinfección esté presente. En ese caso, la lesión aún sería sólo microscópica. En esa situación, la prueba directa, de confirmación de la validez de la prueba indirecta, está expuesta a no proporcionar seguridad total sobre la presencia de infección tuberculosa (4).

#### PROBABILIDAD QUE UN BOVINO POSITIVO A LA TUBERCULINA ESTE REALMENTE INFECTADO

A manera de ejemplo supóngase que en una región dada el servicio veterinario está sometiendo a prueba periódicamente una masa de 100.000 bovinos lecheros y se desea saber cuál es la proporción (P) de animales que siendo positivos a la prueba (I+) son realmente infectados (D+) con *M. bovis*, esto es

$$P^* (D+/I+)$$

o sea, se está preocupando con la eficacia diagnóstica de la prueba de tuberculina. Siguiendo con el ejemplo dado (Cuadro 2), y aplicando el teorema de Bayes (2):

$$P(D+/I+) = \frac{P(D+) P(I+/D+)}{P(D+) P(I+/D+) + P(D-) P(I+/D-)}$$

se puede evaluar la eficacia diagnóstica de la prueba de tuberculina, que para dicho ejemplo sería de

\* P = Probabilidad

$$P(D+/I+) = \frac{(0,01) (0,95)}{(0,01) (0,95) + (0,99) (0,1)} = 0,08756$$

CUADRO 2. Distribución de frecuencias para prevalencia de 1%, sensibilidad de 95% y especificidad de 90%

Situación prueba	Situación real (directa)		
	D+	D-	Total
I+	950	9.900	10.850
I-	50	89.100	89.150
Total	1.000	99.000	100.000

Resulta así que sólo el 8,8% de los positivos a la prueba, en este ejemplo, sería realmente ganado infectado de tuberculosis. No fue utilizada directamente la razón  $950/10.850 = 0,088$  porque, por regla general, no se conoce la cantidad de animales en la célula "a" del Cuadro 1, que es 950 de acuerdo con los datos del ejemplo. Debido a esto hay que utilizar los datos marginales para estimar la eficacia diagnóstica de una prueba indirecta como es la tuberculina (5).

Siendo la prevalencia de infección de 1% y el tamaño de la población de  $n = 100.000$  bovinos, entonces el número de animales positivos o infectados es de 1.000, que corresponde en el Cuadro 2 al total de la primera columna (a+c), o sea el total de verdaderamente positivos [(a+c) = 1%(n)]. Dado que la sensibilidad de la prueba indirecta es de 95%, a través de ella será posible detectar 950 animales de los 1.000 infectados [a = 95% (a+c)].

Este valor 950 corresponde en el Cuadro 2 a la célula "a" de animales realmente positivos detectados también por la prueba diagnóstica indirecta. La diferencia  $1.000 - 950 = 50$  constituye la célula "c" que corresponde a animales realmente positivos no detectados por la prueba indirecta [c = {(a+c) - a}]. Esta cantidad representa los falsos negativos. La segunda columna del Cuadro 2 corresponde a los realmente negativos, cuyo total (b+d) se obtiene por diferencia entre el tamaño de la población (n) y la cantidad total de individuos positivos (a+c), o sea [n - (a+c) = (b+d)]. La cantidad (b+d) es igual a  $100.000 - 1.000 = 99.000$  bovinos negativos.

Puesto que la especificidad de la prueba diagnóstica indirecta es de 90%, mediante ella es posible detectar 89.100 animales de los 99.000 negativos [ $d = 90\% (b+d)$ ]. Esa cantidad corresponde a la célula "d" del Cuadro 2, que representa los bovinos realmente negativos, identificados también como tales por la prueba indirecta. La diferencia  $99.000 - 89.100 = 9.900$  constituye la célula "b", que corresponde a los animales realmente negativos no identificados como tales por la prueba indirecta [ $b = \{(b+d) - d\}$ ]. Esta cantidad representa los falsos positivos.

El total de animales de la primera fila (a+b) corresponde al total de infectados, clasificados como tales por la prueba diagnóstica indirecta, al paso que el total de la segunda fila (c+d) representa el total de no infectados clasificados en esa categoría por la prueba diagnóstica indirecta. En estos últimos dos totales hay una cierta magnitud de errores de clasificación, porque el primer total (a+b) incluye falsos positivos y el segundo (c+d) total incluye falsos negativos.

#### VALOR PREDICTIVO DE LA PRUEBA DIAGNOSTICA

Para estudiar todos los elementos que pueden afectar los resultados de la aplicación de una prueba diagnóstica masiva es conveniente que se consideren algunas situaciones en que la sensibilidad y especificidad de una prueba se mantienen constantes, haciendo variar la prevalencia de infección. Supóngase que se aplica un test diagnóstico indirecto, como la prueba tuberculínica a tres poblaciones de 100.000 bovinos cada una.

La sensibilidad y la especificidad se toman como 91% y 89% respectivamente. La primera población tiene una prevalencia de 1%, la segunda de 10% y la tercera de 20%. Los resultados se muestran en los Cuadros 3, 4 y 5.

Existen dos indicadores de la capacidad predictiva de un test indirecto (6), el valor predictivo positivo (G) y el valor predictivo negativo (H), cuyas expresiones se elaboran a partir de las frecuencias del Cuadro 1.:

$$G = a/(a+b)$$

$$H = d/(c+d)$$

Para los datos del Cuadro 3 los indicadores alcanzan los siguientes valores:

$$G = \frac{910}{11.800} \times 100 = 7,8\%$$

$$H = \frac{88.110}{88.200} \times 100 = 99,9\%$$

CUADRO 3. Población bovina con prevalencia de 1%, sensibilidad 91% y especificidad 89%

	D+	D-	Total
I+	910	10.890	11.800
I-	90	88.110	88.200
Total	1.000	99.000	100.000

Para los datos del Cuadro 4 los indicadores de capacidad predictiva de la prueba tuberculínica alcanzan los siguientes valores:

$$G = \frac{9.100}{19.000} \times 100 = 47,9\%$$

$$H = \frac{80.100}{81.000} \times 100 = 98,9\%$$

CUADRO 4. Población bovina con prevalencia de 10%, sensibilidad 91% y especificidad 89%

	D+	D-	Total
I+	9.100	9.900	19.000
I-	900	80.100	81.000
Total	10.000	90.000	100.000

Como se puede observar, el cambio de prevalencia de +9 puntos porcentuales produjo una modificación de la capacidad predictiva positiva de +40 puntos porcentuales y -1 punto porcentual en la capacidad predictiva negativa. Este último aspecto prácticamente no se modificó cuando es comparado con el incremento experimentado por la capacidad predictiva positiva.

Para los datos del Cuadro 5 los indicadores presentan los siguientes valores:

$$G = \frac{18.200}{27.000} \times 100 = 67,4\%$$

$$H = \frac{71.200}{73.000} \times 100 = 97,5\%$$

CUADRO 5. Población bovina con prevalencia de 20%, sensibilidad 91% y especificidad 89%

	D+	D-	Total
I+	18.200	8.800	27.000
I-	1.800	71.200	73.000
Total	20.000	80.000	100.000

Nuevamente se observa que cuando ocurre un aumento marcado en la prevalencia, se produce también un incremento notable de la capacidad predictiva positiva (prácticamente 20 puntos porcentuales con relación a la situación cuando la prevalencia era 10%). Sin embargo, el crecimiento experimentado por este indicador no tuvo la magnitud del observado cuando la prevalencia pasó de 1% a 10%. Por otro lado, también nuevamente H disminuyó levemente, pero la magnitud de cambio fue en este caso algo mayor que cuando la prevalencia pasó de 1% a 10%.

Estos ejemplos permiten concluir que cuando se utiliza una prueba diagnóstica indirecta, su capacidad predictiva positiva es disminuida si la población bovina tiene un nivel de infección muy bajo.

De acuerdo con Ranney (4), la aplicación de estos resultados a un programa de control y erradicación de la tuberculosis bovina indicaría que en situaciones de baja prevalencia, ya sea por la propia historia natural de la enfermedad o por efecto de acciones de control, el empleo de la prueba tuberculínica como único método de detección de infección no es aconsejable. En estos casos debe haber una actividad de detección dinámica haciendo uso de procedimientos que permitan un rastreo desde el matadero (hallazgo de lesiones) hasta el rebaño, donde hay actividad del agente. Identificados los rebaños problemas, la prueba tuberculínica puede ser aplicada cada 6 meses hasta la limpieza del rebaño.

En situaciones epidemiológicas en las que la prevalencia es muy alta, el valor predictivo positivo de la prueba tuberculínica es también muy alto, de ahí que el método de prueba tuberculínica y eliminación de reactores positivos puede dar buenos resultados.

#### ESTIMACION DE LA PREVALENCIA REAL DE UNA ENFERMEDAD COMO LA TUBERCULOSIS A PARTIR DE LOS RESULTADOS DE LA PRUEBA INDIRECTA DE LA TUBERCULINA

En el Cuadro 6 aparece una posible situación de la tuberculosis en una región.

CUADRO 6. Rebaño bovino con prevalencia 1/1000, sensibilidad 91% y especificidad 89%

Situación prueba indirecta	Situación real (directa)		
	D+	D-	Total
I+	91	10.989	11.080
I-	9	88.911	88.920
Total	100	99.900	100.000

En este caso la prevalencia aparente (PAP) es:  $11.080/100.000 = 0,11080$ , es decir, un 11% aproximadamente. Sin embargo, de los 11.080 positivos a la prueba indirecta, 10.989 son falsos positivos. Este resultado debería ser interpretado con cuidado considerando las implicaciones económicas ocasionadas por el envío al matadero (sacrificados) de reactores positivos a la prueba indirecta, como es en el caso de la tuberculosis (5). El conocimiento que se tenga de los indicadores de sensibilidad y especificidad de la prueba ayuda a resolver este problema. De esta manera se puede llegar a conocer la verdadera tasa de prevalencia (VEP):

$$VEP = \frac{PAP + V - 1}{V + U - 1} = \frac{0,00080}{0,80000} = 0,001$$

Al observar los datos se confirma que la verdadera prevalencia es 1 por 1.000.

#### PROBABILIDAD DE FALLAR EN DETECTAR ANIMALES INFECTADOS, UTILIZANDO UNA PRUEBA INDIRECTA, COMO LA TUBERCULINA

Se desea llevar adelante una serie de diagnósticos para detectar la tuberculosis en un rebaño de animales con una prueba indirecta con sensibilidad y especificidad conocidas. Se quiere conocer la probabilidad de fallar en detectar animales

infectados por el *M. bovis*, es decir, se quiere conocer la probabilidad de obtener falsos negativos. Las variables que participan en este problema son:

U, V y VEP

Despejando la fórmula de la sección anterior se tiene:

$$VEP (V + U - 1) - V + 1 = PAP$$

cuya aplicación en el ejemplo dado últimamente es:

$$PAP = 0,001 (0,89 + 0,91 - 1) - 0,89 + 1 = 0,11080$$

$$PAP = p = 0,11080$$

$$(1 - PAP) = q = 1 - p = 1 - 0,11080 = 0,88920$$

En estas condiciones es posible aplicar la ecuación de la distribución geométrica para evaluar la probabilidad de fallar en la detección de positivos con la prueba indirecta, a la probabilidad de obtención de falsos negativos (7).

$$\left\{ \begin{array}{l} \text{Probabilidad fallar} \\ \text{detectar positivos} \end{array} \right\} = Pr(\text{Fallar} +) = \alpha$$

$$= P(\text{Fallar} +) = (1-p)^n = q^n = \alpha$$

$$\alpha = q^n$$

$$\log \alpha = n \log q \quad \text{donde } n = \frac{\log \alpha}{\log q}$$

Este es el número de individuos necesarios o tamaño mínimo de la muestra para tener una probabilidad igual o menor que " $\alpha$ " de fallar en la detección de positivos con la prueba indirecta, es decir, la probabilidad de obtener falsos negativos. Esta ecuación es aplicable cuando se desea hacer un estudio para detectar la presencia de individuos positivos y sólo es posible hacerlo por muestreo. En tal caso teniendo un " $p$ " hipotético muy bajo, " $q$ " se acerca a la unidad y en términos logarítmicos el denominador de la ecuación se aproxima a cero. Por otra parte, tratándose de verificar si una zona es "limpia" de una enfermedad, entonces " $\alpha$ " tendrá que ser muy pequeño, puesto que se trata de la probabilidad de fallar en detectar positivos, razón por la cual el numerador de la fórmula se tornará cada vez más pequeño en la escala logarítmica (-3, -5, -6, etc.) con la cual, el número mínimo de observaciones en la muestra irá siendo así mayor.

## DISCUSION

En salud animal es frecuente que la detección de infección o enfermedad en un animal no sea establecida con seguridad. Los procedimientos diagnósticos de alta confiabilidad si bien son deseables no es menos cierto que con frecuencia son impracticables, como en la tuberculosis, o son muy costosos o muy laboriosos para ser aplicados masivamente en poblaciones ganaderas.

Al estudiar la situación epidemiológica de una enfermedad en una población animal de cierta magnitud, es común tener que aplicar alguna prueba clasificatoria de positivos no muy sofisticada y que generalmente tiene un nivel de error de clasificación del estado epidemiológico de los individuos.

En cualquier caso, para facilitar la interpretación de los resultados, al aplicar una prueba indirecta tipo clasificatoria es necesario evaluar su validez tomando como referencia las respuestas dadas a una prueba directa. Los niveles de error, falsos positivos y falsos negativos pueden ser estimados específicamente cuando la prueba clasificatoria es aplicada a un grupo de animales cuyo estado epidemiológico verdadero es conocido.

La sensibilidad y la especificidad, relacionadas con los falsos negativos y los falsos positivos respectivamente, son valores característicos propios de una prueba, no dependiendo de la prevalencia de la enfermedad. Sin embargo, los valores predictivos no son estables y cambian de acuerdo con la prevalencia. Cuando baja la prevalencia de una enfermedad, el valor predictivo positivo también baja, lo que lo convierte en una medida no confiable.

En el caso de algunas enfermedades crónicas que afectan al ganado, como es el caso de la tuberculosis bovina, la aplicación de pruebas clasificatorias está obviamente siempre relacionada con su control y erradicación. De ahí que se deba dar atención a algunos aspectos señalados en este trabajo.

En los programas de combate a la tuberculosis, una preocupación puede ser prevenir su ingreso en regiones o grupos de rebaños libres del *M. bovis*. En este caso, una prueba de alta sensibilidad es

especialmente necesaria ya que interesa que no se "pase" ningún falso negativo. La especificidad en estas circunstancias juega un rol secundario.

Otra situación que se puede presentar en los programas de control de enfermedades como la tuberculosis bovina es la necesidad de eliminar animales infectados desde una región o desde un grupo de rebaños. También en esta circunstancia se requiere en forma primordial una prueba con alta sensibilidad, ya que de otra manera el programa estaría expuesto a que muchos animales infectados no fueran detectados, permaneciendo en los rebaños y difundiendo la enfermedad.

Otra situación diferente que podrían enfrentar los programas de combate a la tuberculosis es el acompañamiento de áreas libres de la enfermedad para detectar su presencia. En ese caso la especificidad tiene más importancia que la sensibilidad. Es común que en una región libre, bien protegida por el programa, cuando aparecen aisladamente animales reactivos se trate de falsos positivos.

En resumen, una buena prueba clasificatoria es aquella en que las distribuciones de frecuencias de positivos y negativos tienen la menor área de superposición posible. De esta manera, la prueba

tendrá una alta sensibilidad y especificidad simultáneamente. Definida el área de superposición de las distribuciones de frecuencia en una prueba indirecta, cualquier intento de aumentar su sensibilidad se hará en desmedro de la especificidad y vice versa.

#### REFERENCIAS

1. FELLER, W. An introduction to probability theory and its applications. Vol. 1. Wiley. New York. 1968.
2. FLEISS, J.L. Statistical Methods for Rates and Proportions. Wiley. New York. 1973.
3. FRANCIS, J., SEILER, R.J., WILKIE, I.W., O'BYLE, D., LUMSDEN, M.J., FROST, A.J. The sensitivity and specificity of various tuberculin tests using bovine PPD and other tuberculins. *Vet. Rec.* 103: 420-435, 1978.
4. RANNEY, A.F. Los servicios de inspección de carne como ayuda en la erradicación de la tuberculosis. I Seminario Internacional sobre Tuberculosis bovina para las Américas. Santiago, Chile. OPS, *Pub. Cient.* 258, 1972.
5. SCHWABE, C.W., RIEMANN, H.P., FRANTI, C.E. Epidemiology in Veterinary Practice. Lea & Febiger. Philadelphia, 1977.
6. TOMAN, K. Sensibilidad, especificidad y valor predictivo de los tests diagnósticos. Boletín Unión Internacional contra la Tuberculosis 56, No. 1-2, 1981.