

VET

Héctor Tarabla
Marcelo Signorini

EPIDEMIOLOGÍA DIAGNÓSTICA

EPIDEMIOLOGÍA
DIAGNÓSTICA

UNIVERSIDAD NACIONAL
DEL LITORAL



Epidemiología diagnóstica





**UNIVERSIDAD
NACIONAL
DEL LITORAL**

Rector **Enrique Mammarella**

Director de Planeamiento y Gestión Académica **Daniel Comba**

Directora Ediciones UNL **Ivana Tosti**

.....

Tarabla, Héctor
Epidemiología diagnóstica / Héctor Tarabla;
Marcelo Signorini. - 1a ed. - Santa Fe :
Ediciones UNL, 2020.
Libro digital, PDF - (Cátedra)

Archivo Digital: descarga y online
ISBN 978-987-749-209-5

1. Medicina Veterinaria. 2. Epidemiología. I.
Signorini, Marcelo. II. Título.
CDD 636.08944

.....

© Héctor Tarabla, Marcelo Signorini, 2020.

© ediciones  UNL, 2020

Coordinación editorial
María Alejandra Sedrán
Coordinación diseño
Alina Hill
Producción general
Ediciones UNL

—
editorial@unl.edu.ar
www.unl.edu.ar/editorial

.....



hdl.handle.net/11185/5548

Epidemiología diagnóstica

Héctor Tarabla
Marcelo Signorini

Índice

7	Prólogo a la 1º edición
9	Prólogo a la 2º edición
	Capítulo 1
15	Historia
	Capítulo 2
19	Pruebas diagnósticas
	Capítulo 3
25	Características de las pruebas
	Capítulo 4
39	Relación entre el resultado de la prueba y el estado de salud de los individuos
	Capítulo 5
43	Sensibilidad y especificidad
	Capítulo 6
51	Determinación de la sensibilidad y la especificidad de la prueba
	Capítulo 7
77	Prevalencia real y prevalencia aparente
	Capítulo 8
81	Valor predictivo
	Capítulo 9
101	Eficiencia
	Capítulo 10
107	Combinación de pruebas diagnósticas
	Capítulo 11
131	Poder discriminatorio de la prueba diagnóstica y prevalencia de la enfermedad
	Capítulo 12
141	Interpretación de pruebas diagnósticas a nivel de grupos

Capítulo 13

153 **Aproximación probabilística a problemas
con las pruebas diagnósticas**

Capítulo 14

161 **Elección de la prueba**

Prólogo a la primera edición

En las ciencias médicas, los estudiantes aprenden a lo largo de su carrera una serie de agentes patógenos y profundizan acerca de las diversas patologías, síntomas y signos que producen cada uno de ellos sobre los animales y el hombre. Sin embargo, luego de recibidos se ven enfrentados con una realidad exactamente opuesta. En la práctica diaria enumeran en sus pacientes una serie de signos, síntomas y resultados de pruebas diagnósticas y deben dilucidar cuál es o cuáles son sus causas más probables. Igualmente, los planificadores de programas de salud rara vez conocen la frecuencia real de la enfermedad en la población. Por lo contrario, toman sus decisiones de acuerdo con los resultados obtenidos en diversas maniobras diagnósticas practicadas y cuantifican los individuos aparentemente enfermos.

En la actualidad existe una gran variedad de pruebas diagnósticas que son utilizadas para diferenciar animales o grupos

de animales supuestamente sanos de enfermos, estimar la prevalencia de enfermedades, discriminar individuos enfermos para someterlos a tratamientos específicos o evitar el ingreso de hospedadores infectados a regiones con programas de control o erradicación de la afección de interés. Sin embargo, los factores subyacentes que influyen los resultados de la prueba no siempre han sido correctamente identificados y comprendidos por los encargados de la toma de decisiones, tanto a nivel de planificación, de laboratorio como de campo. A lo largo de este texto se presentarán los conceptos básicos sobre el tema, sin profundizar en su desarrollo algebraico o probabilístico. Las probabilidades se presentarán en forma de porcentajes para simplificar la discusión. La intención del autor es ofrecer un texto que prefiera los aspectos prácticos con ejemplos que faciliten al lector la comprensión de los problemas reales.

Prólogo a la segunda edición

Cuando el Dr. Tarabla y el Dr. Signorini me pidieron que escribiera el Prólogo a la segunda edición de *Epidemiología diagnóstica* quizás no aquilataron el honor que me hacían, ya que a pesar de que mi vida académica y profesional la he dedicado a esta temática, aún no he podido entregar mis vivencias en un libro, como ellos lo están haciendo. Con esto quiero decir lo difícil y dedicado que es hacer un libro y, en particular, para apoyar el quehacer de los epidemiólogos y estudiantes de ciencias veterinarias; pero que también podrán usar muchos otros profesionales de los fundamentos que nos entregan en sus líneas y que tanta falta hace en nuestros países de habla hispana.

Quisiera señalar que hay tantas definiciones de epidemiología como autores han escrito sobre el tema, y el énfasis de todas ellas está orientado hacia el sujeto de acción, ya sea que se trate del ser humano o de los animales. En el caso de *Epidemiología diagnóstica* el foco se orienta a las especies animales productivas terrestres; pero también lo podemos aplicar a especies de compañía, animales acuáticos y otras. Hay elementos en todas esas definiciones en las que coinciden la gran mayoría de ellas, y son que a través de la Epidemiología se estudia la salud y enfermedad de las poblaciones, y la finalidad de todos sus

procedimientos y estrategias es mantener la salud y prevenir, controlar o erradicar la enfermedad de esas poblaciones. En mi experiencia, y como aporte al libro de mis amigos, me atrevo a decir que para mí la epidemiología es “una forma de pensar, para fomentar y mantener los patrones de salud y prevenir, controlar o erradicar los patrones de enfermedad, en poblaciones animales”.

La Epidemiología, vinculada con la enfermedad de los animales, se remonta a la historia del hombre ya que, aunque no se la describiera con ese nombre, lo que practicaban eran procedimientos que hoy se ven claramente incorporados en el quehacer epidemiológico. Quienes realizaban esas funciones no eran los hoy en día médicos veterinarios, sino que pertenecían a las más diferentes actividades de las sociedades de aquellas épocas, como curanderos, herreros, brujos y, en la mayoría de los casos, personas interesadas en el problema de salud pública o salud animal.

Hasta la segunda mitad del siglo XX, las explicaciones a las distintas enfermedades eran casi completamente debidas a los agentes y sus efectos en los hospederos, pero en un momento dejaron de ser suficientes y fue necesario incorporar los factores ambientales y las políticas sanitarias. Es en esta etapa de la historia cuando

comienzan a desarrollarse con mucha fuerza los diagnósticos epidemiológicos, con una visión holística e incorporando las descripciones de enfermedad en su dimensión temporal y espacial, los análisis cualitativos y cuantitativos, los sistemas de vigilancia y monitoreo y el trabajo en detalle con las pruebas diagnósticas, los elementos que nos ofrecen Tarabla y Signorini en su libro.

La búsqueda de la enfermedad, sin embargo, nos obliga a mirar también a los animales que están sanos, porque debemos responder la pregunta ¿para qué se desea tener la condición de salud y por qué se debe trabajar para mantenerla? Ello obliga a incorporar la mirada económica, social y política que tiene la epidemiología, en las especies animales de importancia económica.

A modo de ejemplo, tener sin tuberculosis un rebaño lechero interesa sobremanera, porque se desea entregar un producto de calidad sanitaria; pero si la productividad de ese rebaño es un litro por vaca se estará en problemas para mantener el negocio. De ahí que, con mucha propiedad, se pueda decir que ese “sistema productivo se encuentra enfermo”, ya que no se está obteniendo lo esperado, aun cuando no tenga enfermedades producidas por agentes biológicos específicos.

Quisiera en esta oportunidad, y aprovechando la gentileza de mis amigos, recordar a una de las personas que aportó a clarificar muchos conceptos en nuestro trabajo epidemiológico al definir el concepto de salud animal y al adaptar herramientas de trabajo en proyectos de planificación, para el control de enfermedades animales; es el Ing. Hidráulico Osvaldo Fernández Balma-ceda quien trabajó en la Oficina Paname-

ricana de la Salud. El Ingeniero nos dijo que salud animal es “un estado óptimo de productividad”, sin procesos mórbidos y sin afectar al ser humano, entre otros alcances. Dicho concepto comprende que la salud es un estado, es decir, la condición cambiante y, por lo tanto, es dinámica en el tiempo. Debe, además, tender a la optimización, es decir, obtener el máximo posible de acuerdo con los recursos disponibles; finalmente, la productividad quiere decir que ese máximo es por unidad de insumo que ponemos en el sistema. Por lo anterior, la finalidad de la salud no es solamente no tener enfermedades, sino expresar en forma óptima las potencialidades genéticas, según los aportes entregados.

Se entiende entonces que el quehacer epidemiológico debe orientarse a mantener la salud, para expresar al máximo las potencialidades genéticas en las mejores condiciones de bienestar posibles y al mismo tiempo prevenir, controlar y erradicar todas las enfermedades, sean biológicas, físicas, químicas, de manejo u otras causas. Es entonces adecuado pensar que como epidemiólogos estamos al servicio de la producción animal, las formas de producción y de sus cambios que dependen, en gran medida, del mercado.

Por otra parte, todos quienes han definido epidemiología concuerdan en que se trata de acciones que se realizan sobre poblaciones, lo cual aparece claramente expresado en los diferentes capítulos del libro. Esto da al médico veterinario una nueva forma de mirar la salud y la enfermedad desde las caracterizaciones poblacionales: de las migraciones de animales silvestres a las de comercialización de ellos,

donde aparecen las diferentes estrategias de trazabilidad.

Lo expuesto ha significado un crecimiento enorme de la epidemiología cuantitativa, lo cual se puede apreciar en la inmensa cantidad de publicaciones que incorporan elementos estadístico-poblacionales; tanto es así, que muchas veces quienes trabajan hace muchos años en epidemiología se preguntan si lo que están haciendo es estadística o epidemiología.

A mi entender, lo que más ha empujado al desarrollo de la forma de pensar y el trabajo epidemiológico, en su dimensión poblacional, han sido la herramienta computacional y el apoyo con programas extraordinariamente eficientes que permiten trabajar con muchos individuos al mismo tiempo y con una cantidad enorme de variables, lo que hace algunos años era impensable. Para el tema del libro también tenemos a la mano programas computacionales que nos ayudan a trabajar las muestras de animales a diagnosticar, con las sensibilidades y especificidades de las pruebas y comprender las prevalencias aparentes y reales de los grupos.

La economía ha sido una de las últimas herramientas incorporadas en los análisis epidemiológicos, debido a las exigencias de la banca para prestar dinero para proyectos destinados a controlar o erradicar enfermedades de los animales directamente o que afectaban su comercio o productos. Esto se ha sentido desde las asesorías técnicas que ofrecen los médicos veterinarios a predios individuales hasta los proyectos que los países desean llevar adelante en sus territorios. Esta problemática se expresado, en sus diferentes dimensiones, en los Simpo-

sios Internacionales de Epidemiología y Economía Veterinaria realizados cada tres años; el XIII se realizó en agosto de 2012 en Maastricht, Holanda, y el X, en noviembre de 2003, en Latinoamérica, en Viña del Mar, Chile, donde los autores de este libro tuvieron una destacada participación.

Quisiera expresar unas palabras sobre la producción de los animales y su medición económica, que son también formas de establecer condiciones de salud de los animales. Han aparecido diseños de monitoreo de grupos animales que les miden diferentes parámetros sanitarios y productivos y se comparan con otros de condiciones similares, para apreciar sus fortalezas y debilidades. Así es posible apreciar la necesidad de mejorar o incorporar en el área de la producción animal elementos nutritivos nuevos o combinaciones genéticas, tecnologías de manejo reproductivo y otras acciones de manejo, tendientes a controlar o prevenir la ocurrencia de enfermedades biológicas o deficiencias de algún tipo. Con ello se le agrega mayor valor a la producción animal, mediante la agroindustrialización, diversificación de la producción, perfeccionamiento de los mercados internos, administración de la oferta, etc., todas demandas del mercado que, a su vez, hacen cambiar las estrategias para la salud animal.

La globalización del mundo y en particular la comercialización de animales y productos han establecido una interdependencia entre países y continentes tal que establece una dimensión nueva de trabajo para la epidemiología, con el concepto de *One Health*. Además, dentro de los países se ha obligado al acercamiento de trabajo entre el sector público y el sector privado.

Era normal pensar que el Estado, que velaba por el bien común, fuese el responsable por las acciones que vigilaban estrechamente al sector privado, para que cumpliera con los procedimientos establecidos, fundamentalmente con los mercados internos. Pero al abrirse las fronteras, el bien común se ha transformado; es importante ganar mercados con productos de calidad e inocuidad requeridos por la demanda y, por lo tanto, acercar los propósitos de ambos sectores. De ahí que no es extraño ver trabajar conjuntamente al Estado y al sector privado, para mejorar las condiciones que le permitan al país ganar un mercado.

Es por ello que el Estado ha estado cambiando su cultura y sus profesionales se están adaptando a una nueva forma de trabajo que abandona la acción policial por una de mayor colaboración y apoyo a los productores privados. Por otra parte, la globalización del intercambio de animales y sus productos ha establecido que los estudios de análisis de riesgo estén cobrando gran importancia, ya que no es aceptable que se establezcan las condiciones de comercialización unilateralmente. Así, en el área de la epidemiología estatal se está trabajando con mucha fuerza en estos temas, para que la salud animal no se transforme en una "barrera no arancelaria" que impida el comercio honesto y abierto entre los países.

Cabe destacar que los aspectos legales de los intercambios comerciales internacionales surgen con mayor frecuencia y los epidemiólogos estamos llamados a ofrecer nuestras capacidades científicas y tecnológicas en esos litigios legales, que en muchos casos tienen relación con acusaciones de *dumping* o de prácticas comerciales inadecuadas.

También, el consumidor está empezando a jugar un papel mucho más activo al hacer respetar sus derechos a comer productos sanos y de calidad, y exigir en muchos casos certificaciones de origen, formas de producción de acuerdo con sus principios éticos, formas de faenamiento y bienestar de los animales, como parte del pago que él hace por los productos. Esto debe también repercutir en las universidades y sus programas de estudio, para que los estudiantes de epidemiología tengan los elementos necesarios que les permitan trabajar y servir.

La presente edición de Tarabla y Signorini nos acerca un conjunto secuencial de temas que ayudan a enfrentar en forma sencilla, pero eficaz, el diagnóstico de las enfermedades, aspecto crucial para entender bien lo que está pasando en la sanidad de nuestras especies animales. Con esa plataforma conocida se podrán entregar opciones estratégicas que permitan acciones sobre esas enfermedades, que limitan posibilidades económicas de negocios de exportación o afecten a la salud pública de los habitantes de nuestros países.

Así como la Epidemiología en el campo de las Ciencias Veterinarias ha tenido una importante evolución, no encontramos mucho apoyo en la literatura en español cercana al ejercicio de la práctica veterinaria en el área, como este libro que tenemos la oportunidad de tener con nosotros.

Por eso, al mismo tiempo que observamos grandes progresos en la modelación de enfermedades, con herramientas computacionales avanzadas y análisis de riesgos con múltiples variables, no sabemos exactamente los errores que pueden tener los antecedentes empleados para la realización de esos análisis. También, las simulaciones

y evaluaciones económicas requieren sustentarse en buenos diagnósticos, con pruebas estandarizadas y cuyas bondades y debilidades podamos comprender. De esa manera lograremos hacer buenas interpretaciones y que los progresos esperados en lo sanitario estén fuertemente cimentados en óptimas pruebas diagnósticas y con profesionales que sepan emplearlas, interpretarlas y hacerlas comprender a todos los agentes que tienen que tomar decisiones políticas y económicas.

Este libro está escrito por dos profesores e investigadores que han desarrollado su trabajo en el Instituto Nacional de Tecnología

Agropecuaria, una de las instituciones más señeras en la labor de conocer la realidad agropecuaria de la República Argentina. Por ello, se tiene una especial garantía respecto de que los conceptos vertidos se fundamentan en una experiencia largamente trabajada, en condiciones de campo y con realidades socioeconómicas y culturales diversas.

Con especial agradecimiento y aprecio por este regalo a la Epidemiología Veterinaria, saludo con gran aprecio a Héctor y Marcelo, deseándoles mucho éxito con esta segunda edición del libro *Epidemiología diagnóstica*.

Santiago Urcelay Vicente

MV; MSP; MPVM

Decano Facultad de
Ciencias Veterinarias y Pecuarias
Universidad de Chile

Capítulo 1

Historia

Los relevamientos efectuados para determinar la prevalencia de las enfermedades en la población y su aplicación en la vigilancia epidemiológica de programas de erradicación fueron primero aplicados en medicina veterinaria en la segunda mitad del siglo XIX y comienzos del siglo XX. En los albores de los servicios veterinarios se carecía de información sobre las características vitales de las poblaciones animales, datos que aún en nuestros días son difíciles de obtener en condiciones de explotación pastoril extensiva. Como consecuencia, los registros de datos de incidencia de enfermedades siempre fueron infrecuentes, por lo que los servicios veterinarios debieron intentar otros enfoques para cuantificar la morbilidad causada por diferentes patologías en las poblaciones animales.

En 1857 James Simonds y Williams Ernes iniciaron en Europa Continental una serie de trabajos, basados mayoritariamente en entrevistas personales con veterinarios, ganaderos y vendedores de hacienda, tratando de justipreciar el riesgo de ingreso de la peste bovina y su impacto en las Islas Británicas (Pattison, 1973:207). Estos trabajos significaron los primeros pasos de la epidemiología y la medicina veterinaria preventiva en Gran Bretaña. La posterior utilización de los relevamientos masivos mediante exámenes clínicos (pleuroneumonía contagiosa en EE. UU.), pruebas intradérmicas (tuberculosis bovina en Europa y EE. UU.) o pruebas serológicas (durina equina en EE. UU.) fueron fundamentales en algunas campañas. El advenimiento de las pruebas serológicas permitió procesar en el laboratorio un gran número de muestras por día. Con esto se tuvo una nueva herramienta de vigilancia epidemiológica de las enfermedades del ganado que dio lugar a la llamada Epidemiología Serológica. Desde entonces, las mediciones periódicas de la prevalencia utilizando técnicas inmunodiagnósticas han sido la espina dorsal de la vigilancia epidemiológica de muchos programas de control de enfermedades en Medicina Veterinaria.

Como contrapartida, estos esfuerzos en gran escala recién comenzaron en medicina humana alrededor de 1920 de este siglo con el uso de la Prueba de Schick para determinar patrones de inmunidad de la difteria en niños o el requerimiento rutinario de la Prueba serológica de Wassermann para el diagnóstico de sífilis en pacientes hospitalarios (Paul, 1973:218). No obstante, el nombre de Epidemiología Serológica se debe al Dr. J. R. Paul (1893–1971), uno de los investigadores que más trabajó en el desarrollo de los relevamientos para determinar prevalencia como una metodología epidemiológica en salud pública (White, 1973:218), un área tradicionalmente dominada por los informes de incidencia como herramienta de vigilancia. Epidemiología serológica proviene del hecho de que la mayoría de las pruebas diagnósticas utilizadas en dichos estudios fueron pruebas realizadas sobre suero sanguíneo. Sin embargo, los principios y conclusiones aquí enunciadas son igualmente aplicables a toda prueba utilizada en la detección de enfermedades en individuos o en poblaciones animales independientemente de su complejidad analítica. Por ejemplo, el sistema tradicional de inspección *post-mortem* en frigoríficos y mataderos que aún está vigente está basado en la observación, la palpación y los cortes con cuchillos. Este método diagnóstico fue desarrollado en Europa a mediados de 1880 para detectar enfermedades endémicas (tuberculosis, triquinosis y teniasis) (Blackmore, 1986), poco antes de que se demostrara que la tuberculosis y la brucelosis podían ser transmitidas de los animales al hombre a través de carne y leche infectada, respectivamente (Ostertag, 1899). Para incluir las pruebas diagnósticas no serológicas se propuso en la primera edición el nombre de Epidemiología Diagnóstica.

En el campo de la investigación médica, la estadística clásica ha supuesto y supone un importante aporte de rigor metodológico. No obstante, su utilización rígida y en cierto modo dogmática la aleja en determinadas ocasiones de su verdadera misión como herramienta en las ciencias médicas. Como contraposición a la estadística clásica, los métodos bayesianos aportan una interpretación diferente del concepto de probabilidad. La regla de Bayes (en honor al teólogo y matemático, Reverendo Thomas Bayes, quien fue el primero en utilizar la probabilidad inductivamente y establecer una base matemática para la inferencia probabilística) es una técnica para calcular probabilidades condicionales. A partir de un conjunto de probabilidades llamadas *a priori* o «sin corregir», calcula un conjunto de probabilidades *a posteriori* o «corregidas» que no son más que una modificación de las primeras ante la evidencia de que un determinado suceso ha ocurrido. La aplicación más intuitiva en medicina de este teorema la encontramos en el campo de las pruebas diagnósticas; nos permite, conociendo la prevalencia de una enfermedad en la población a la que pertenece un individuo y los valores de sensibilidad y especificidad de la prueba, calcular la probabilidad de que un sujeto que ha dado positivo a una prueba diagnóstica verdaderamente tenga esa enfermedad. La ventaja del análisis bayesiano al interpretar una prueba diagnóstica consiste en que el clínico logra una mejor estimación del riesgo que tiene un paciente

de tener o contraer una enfermedad cuando la prueba le da positiva. No obstante, aun considerando sus ventajas, la utilización de la metodología bayesiana está muy lejos de ser frecuente.

Las pruebas indirectas tienen actualmente un rol fundamental en el estudio de enfermedades, tanto a nivel individual como poblacional. En este último, su uso ha sido limitado como un procedimiento de vigilancia para hallar casos o grupos de animales afectados de una enfermedad en particular. Por ejemplo, en Argentina y otras partes del mundo se recolectaron grandes cantidades de muestras sanguíneas para diagnóstico de brucelosis por la técnica del antígeno bufferado (BPA). En general, estas muestras han sido rápidamente descartadas sin una utilización posterior, a pesar de su potencial utilidad para (Schwabe et al., 1982:303): a) los relevamientos de otras enfermedades infecciosas; b) las estimaciones de prevalencia de enfermedades metabólicas, carenciales o tóxicas; c) el mapeo de la distribución y frecuencia de marcadores genéticos; d) el uso de poblaciones animales para el monitoreo de radiaciones, pesticidas u otros contaminantes del ambiente; y e) el almacenamiento en bancos de sueros de especímenes seleccionados para futuros estudios.

De igual manera, otros tipos de muestras tomadas en forma rutinaria pueden tener valiosos usos alternativos. Por ej., las muestras de leche tomadas en el monitoreo de células somáticas en rodeos lecheros pueden ser usadas para la detección de antibióticos y otros inhibidores en leche, la detección de nuevos patógenos en un área geográfica, los estudios de prevalencia de infecciones intramamarias y el monitoreo de contaminantes ambientales.

El estudio y comprensión de todos los elementos intervinientes en la interpretación epidemiológica de los resultados obtenidos luego de la utilización de pruebas indirectas en el diagnóstico de enfermedades han ido profundizándose a lo largo de los años. Paralelamente se han ido transfiriendo los conocimientos y la utilidad a nivel de campo de los conceptos teóricos, contando actualmente con programas de computación como el Episcope 2.0 (Frankena et al., 1990), el Testview (Gardner y Holmes, 1993), el PvdT (Greiner, 1998), el MedCalc (Schoonjans et al., 1995), o calculadoras en línea a través de Internet (Kneusel y Barnas, 1995) que facilitan la ejecución de las operaciones necesarias y la presentación gráfica de los resultados. Sin embargo, para el uso óptimo de una prueba diagnóstica siempre se requerirán conocimientos básicos de la patogenia de la enfermedad, el proceso inmunológico y la inferencia estadística.

Bibliografía

- Bayes, T. (1763).** An Essay towards solving a Problem in the Doctrine of Chances. Philosophical Transactions of the Royal Society of London 53: pp. 370–418
- Blackmore, D.K. (1986).** Developments in veterinary public health as they affect meat quality. *Kajian Veterinar Malaysia* 18: 229–234.
- Frankena, K., Noordhuizen, J.P., Goelema, J.O. (1990).** Episcopo: Computer programs in Veterinary Epidemiology. *Vet. Rec.* 126: 573–576.
- Gardner, I.A., Holmes, J.C. (1993).** Testview: A spreadsheet program for evaluation and interpretation of diagnostic tests. *Prev. Vet. Med.* 17: 9–18.
- Greiner, M. (1998).** Pvd. Seroepidemiol Working Group, Inst. parasitol. & Trop. Vet. Med., Dept. Trop. Vet. Med. & Epidemiol. Freie Univ. Berlin, <http://city.vetmed.fu.berlin.de/~mgreiner/pvdt/pvdt.htm>.
- Kneusel, R., Barnas, G. (1995).** Bayesian analysis model. *General Internal Med., Med. Coll. Wisconsin, Milwaukee*, [http://141.106.68.17/bayes.acgi\\$1](http://141.106.68.17/bayes.acgi$1).
- Ostertag. (1899).** The use of flesh and milk of tuberculous animals. *J. Comp. Pathol. & Therapeutics* 12: 240–250.
- Pattison, I. (1973).** The British Veterinary Profession, 1791–1948. J. A: Allen & Co. Ltd., London, 207 pp.
- Paul, J.R. (1973).** Development and use of serum surveys in Epidemiology. En: Paul, J. R. and White, C. 1973. *Serological Epidemiology*. Academic Press, Inc., New York, 218 pp.
- Schwabe, C.W., Riemann, H.P., Franti, C. E. (1982).** *Epidemiology in Veterinary Practice*, 303 pp.
- Schoonjans, F., Zalata, A., Depuydt, C.E., Comhaire, F.H. (1995).** MedCalc: a new computer program for medical statistics. *Comput. Methods & Programs.* 48: 257–262.
- White, C. (1973).** Preface. En: Paul, J.R. and White, C. 1973. *Serological Epidemiology*. Academic Press, Inc., New York, 218 pp.

Capítulo 2

Pruebas diagnósticas

1. Introducción

Una prueba diagnóstica es cualquier proceso diseñado para detectar un signo, una sustancia, un cambio tisular o una respuesta, por medio del cual un individuo, o un grupo de individuos, es clasificado como poseedor o careciente de un determinado atributo (usualmente una enfermedad) (Dohoo et al., 2003:706; Smith, 1991:225). «Diagnóstico» proviene del griego «distinguir, conocer» y se aplica al arte o acto de reconocer una enfermedad (Last, 1983). En este sentido, las pruebas tamiz y las confirmatorias son simplemente etapas para llegar a un conocimiento más certero de la identidad de la enfermedad. Sin embargo, muchas veces el término «prueba diagnóstica» es utilizado como sinónimos de «prueba confirmatoria».

En general, las pruebas diagnósticas son utilizadas para descubrir la presencia de una enfermedad, confirmar su presencia o confirmar su ausencia (Abramson, 1988:326). En este texto se hará referencia a «enfermos» y «sanos», utilizándose estos términos sólo por razones de simplicidad y con respecto a una característica en particular. *Enfermo* se refiere a un individuo poseedor de un atributo que causa una alteración del estado de salud. *Sano* se refiere a otro individuo exento de esa alteración de interés o enfermo de otras alteraciones excluyendo a la afección de interés.

Las pruebas diagnósticas juegan un papel central para la toma de decisiones en la clínica veterinaria. El resultado de una prueba diagnóstica puede ser utilizado para decidir el inicio de un tratamiento, el nivel del tratamiento seleccionado o la evolución del caso clínico posterior al inicio de la terapéutica. Las pruebas diagnósticas son empleadas también a nivel de grupos de animales (rodeo, piara, cardumen, majada, etc.) para determinar la frecuencia de presentación de una enfermedad en la población para identificar la causa o factores predisponentes a una enfermedad o incluso para evaluar el progreso de programas de control o erradicación de enfermedades.

Las pruebas diagnósticas involucran no sólo aquellas desarrolladas en laboratorio, sino también las efectuadas a campo (prueba de mastitis de California), los exámenes físicos, la historia clínica y las imágenes y procedimientos (rayos X y electrocardiogramas). Por otra parte, cabe destacar que las pruebas diagnósticas no son sólo utilizadas para detectar enfermedad, sino también para hallar y/o cuantificar infecciones, infestaciones, intoxicaciones, carencias, perfiles metabólicos, niveles hormonales o estados fisiológicos. Su uso es tan variado que incluye la detección o cuantificación de anticuerpos en suero sanguíneo (ej.: fijación del complemento o aglutinación en tubo en brucelosis), antígenos en raspajes prepuciales (ej.: inmunofluorescencia en campilobacteriosis genital bovina), células somáticas en leche (ej.: prueba de California en mastitis subclínicas) o preñez en animales gestantes (ej.: prueba de la progesterona en leche).

Los resultados de las pruebas diagnósticas pueden ser reportados usando cualquiera de las tres escalas de medición:

- a) Nominal** (dicotómica): signología clínica (presente/ausente), preñez (si/no), aislamiento bacteriano (si/no) o prueba serológica interpretada como positiva/negativa.
- b) Ordinal**: el ejemplo más común corresponde a los títulos serológicos (1:4, 1:8, 1:16, etc.) o la estimación de la condición corporal en vacas lecheras con una escala de medición de 1 (muy flaca) a 5 (muy gorda).
- c) Conteo**: conteo de huevos de parásitos por gramo de materia fecal o glóbulos rojos por mililitro de sangre.
- d) Continuas**: concentración sérica de enzimas hepáticas, milímetros en el diagnóstico intradérmico de tuberculosis bovina.

2. Prueba tamiz

Tamizado es el proceso por el cual en una población se analizan individuos aparentemente sanos para detectar enfermedad. Las pruebas tamices se utilizan en general en enfermedades crónicas o en las que generan respuestas medibles duraderas en el tiempo (ej.: anticuerpos). Se pueden distinguir al menos tres tipos de aplicaciones a nivel poblacional (Last, 1983:114):

- a)** El tamizado masivo que tiene como objetivo detectar una condición en una población numerosa y no seleccionada.
- b)** El tamizado múltiple donde se utilizan varias pruebas tamiz en el mismo relevamiento.
- c)** El hallazgo de casos con el objetivo de detectar precozmente enfermedades en pacientes presuntamente sanos. En estos casos se trabaja sobre un grupo predeterminado para luego derivar los individuos positivos, para ser sometidos a una prueba confirmatoria, a un tratamiento o a su destrucción. Cuando se intenta detectar afecciones poco frecuentes el tamizado suele restringirse a grupos de alto riesgo (Smith, 1991:225).

El tamizado masivo puede ser aplicado a individuos (ej.: tuberculinización intradérmica en el pliegue anocaudal en bovinos) o a grupos de ellos (ej.: digestión enzimática para el diagnóstico de trichinelosis en grupos de porcinos en los frigoríficos, aplicación de la prueba de anillo en leche para detección de rodeos bovinos con brucelosis). El hallazgo de casos se puede realizar, entre otros ejemplos, en individuos que serán introducidos en otra área geográfica que está sujeta a un plan de control o erradicación de la enfermedad (ej.: realización de la prueba ELISA 3 ABC para determinación y valoración de anticuerpos contra proteínas no estructurales del virus de la fiebre aftosa), individuos utilizados como dadores de material biológico (ej.: terneros esplenectomizados utilizados para la producción de vacunas contra anaplasmosis bovina, toros donantes de semen para inseminación artificial, vacas donantes de embriones para trasplante) o individuos utilizados como fuente de alimentos (ej.: inspección de carnes en frigoríficos).

El objetivo de efectuar una prueba tamiz no consiste en realizar un diagnóstico exacto sino solamente identificar individuos o grupos que tengan una alta probabilidad de estar enfermos o poseer algún otro atributo de interés (Abramson, 1988:326). Para que pueda ser empleada masivamente, la prueba tamiz debe ser rápida, barata, segura y poco sofisticada. Se trata solamente de un examen inicial con un grado de error relativamente alto, el cual es indicado generalmente por un investigador o agente de salud y no por el propietario del animal enfermo. Los individuos con resultados positivos a la prueba son usualmente investigados para su mejor diagnóstico y eventuales decisiones subsecuentes (tratamiento, eliminación del rodeo). Un ejemplo típico de prueba tamiz en veterinaria es la prueba de antígeno tamponado o BPA (del inglés *Buffered Plate Antigen*) para el diagnóstico de brucelosis.

3. Prueba confirmatoria

Una prueba confirmatoria es el proceso a través del cual un individuo es clasificado como enfermo o sano. La prueba es solicitada generalmente por el clínico a cargo del caso o bien por los planificadores de un programa de erradicación que requiere la eliminación de los animales positivos. En el primer caso, el juicio del clínico basado en el análisis de los síntomas y signos que presenta el paciente actúa como prueba tamiz. En el segundo caso, los animales con diagnóstico positivo a la prueba tamiz (ej.: prueba de BPA en brucelosis) son luego re-analizados con una prueba más específica (ej.: fijación del complemento). Como se puntualizó anteriormente, el término «prueba diagnóstica» es utilizado muchas veces como sinónimo de «prueba confirmatoria» (Tarabla, 1995).

En algunos casos la diferenciación entre prueba tamiz y confirmatoria no está dada por la prueba en sí sino por la disponibilidad de pruebas diagnósticas y su uso en situaciones reales. Por ejemplo, la prueba de aglutinación en placa (Huddlesson) en

brucelosis o la de inmunodifusión en gel (Coggins) en anemia infecciosa equina fueron utilizadas tanto para monitoreos poblacionales como para el diagnóstico individual de la enfermedad. Por otra parte, la prueba de enzimoimmunoensayo (ELISA) indirecto, donde puede variarse el punto de corte y posee la ventaja de la automatización de la lectura de los resultados, fue recomendada tanto para prueba tamiz como para prueba confirmatoria del diagnóstico de brucelosis (Dohoo et al., 1986). Es por ello que, en lugar de diferenciar la pruebas en sí como tamiz o confirmatoria, en ocasiones resulta más conveniente calificar las estrategias utilizadas.

4. Utilidad de las pruebas diagnósticas

Uno de los problemas de la interpretación de muchas pruebas es que la persona que la efectúa cree que ellas son más útiles de lo que en realidad son (Dohoo et al., 2003:706; Galen y Gambino, 1975:237). Tomemos como ejemplo la prueba VIAA (*Virus Infection Associated Antigen*), basada en la detección de un antígeno asociado a la infección por el virus aftoso. Su aparición creó una expectativa exagerada sobre su utilidad en los programas de control y erradicación de la fiebre aftosa, asignándole en algunas oportunidades el valor de un elemento imprescindible para lograr el control de la enfermedad. Sin embargo, las «encuestas VIA» no medían el grado de protección de la población (ej.: si se había vacunado o no), ni identificaban el tipo de virus actuante, ni servían para evaluar individuos aislados o detectar portadores. Esta prueba demostraba simplemente la presencia o ausencia de los anticuerpos contra el antígeno VIAA en los animales muestreados en la encuesta, e indicaba si se había producido un ciclo activo de multiplicación viral. Con ella sólo se podía estimar la prevalencia y distribución geográfica de la actividad viral en poblaciones de animales. Aún así, estos datos estaban influenciados por factores tales como infecciones pasadas (focos), la presencia de bovinos portadores, las vacunaciones recientes, la edad de los animales, los flujos de movimiento de ganado en la zona y/o el intervalo de tiempo producido desde la infección y la toma de muestras.

Otro ejemplo es el caso de la calcemia en bovinos lecheros como técnica de diagnóstico de la hipocalcemia. Podemos medir el calcio (Ca) en suero sanguíneo con un alto grado de precisión y especificidad química. Podemos obtener excelentes estadísticas relacionando los valores encontrados con factores conocidos que afecten su concentración en sangre como raza, edad y sexo de los animales, dieta a la que están sujetos, en el caso de las hembras su número de partos, estado fisiológico y tiempo de lactancia. Sin embargo, los datos de bioquímica clínica en bovinos enfermos no siguen una distribución gaussiana (Knox et al., 1998). Además, algunos de los factores de la variabilidad estadística de la calcemia no son factores relacionados con la hipocalcemia clínica y su síndrome más frecuente, la paresia posparto. Esta manifestación clínica de la enfermedad puede involucrar cambios importantes en muchos

iones además del Ca (y del magnesio y del fósforo) (Michell, 1992). Por otra parte, se pueden encontrar durante meses niveles disminuidos de Ca en sangre con relación a los estándares en vacas lecheras pertenecientes a diversas categorías sin que se presenten manifestaciones clínicas de hipocalcemia (Marcos, 1982). La «pureza» química y estadística en la medición del Ca sanguíneo no garantiza la utilidad de la prueba en el momento del diagnóstico de las causas por las cuales un animal permanece caído luego del parto y menos aún que el animal responderá positivamente al tratamiento específico.

Como contrapartida, hay casos en que la técnica diagnóstica es en realidad más útil de lo que la persona que la efectúa cree que es. Por ejemplo, el sistema de inspección *post-mortem* en frigoríficos genera un gran volumen de datos de enfermedades tales como la tuberculosis (Abdala et al., 1999) y la hidatidosis (Abdala y Tarabla, 2009), que generalmente no es aprovechado en sistemas de vigilancia epidemiológica. Si se cuenta con un sistema de identificación de los animales práctico y confiable, este diagnóstico permite realizar un seguimiento intensivo de los rodeos positivos, estudiar la evolución temporal y espacial de rodeos afectados y estimar la prevalencia de rodeos enfermos (Pérez et al., 2002; Tarabla y Abdala, 1998).

Entre las aplicaciones más comunes de una prueba diagnóstica se encuentran la obtención de:

- a)** La prevalencia de una enfermedad (ej.: estimación de la prevalencia de infecciones intramamarias).
- b)** La prevalencia de otros atributos (ej.: del grado de protección vacunal en fiebre aftosa por medio de la técnica de seroneutralización, del grado de estabilidad enzoótica de la babesiosis bovina en una área endémica por medio de la inmunofluorescencia indirecta o de la anaplasmosis bovina por medio de la prueba de aglutinación en tarjeta, etc.).
- c)** Una clasificación inicial de rodeos (ej.: prueba de anillo en leche total del tambo para brucelosis bovina para luego efectuar pruebas complementarias individuales en aquellos rodeos positivos a la primer prueba, conteos de células somáticas en leche total del tambo para luego detectar individualmente infecciones intramamarias) o individuos (ej.: Prueba de BPA para brucelosis bovina para luego efectuar la prueba de fijación del complemento en aquellos individuos positivos a la primera prueba) de los cuales se desconoce su estado sanitario con respecto a la enfermedad de interés.
- d)** Un indicio sobre la posibilidad de que una enfermedad esté presente en sus estadios iniciales en un paciente que de otra manera sería clasificado como sano (ej.: radiografías en tumores).
- e)** Un diagnóstico correcto en un paciente del cual se sabe la enfermedad que lo aqueja (ej.: cultivo bacteriano en leche de una vaca con mastitis clínica).

- f) Un soporte adicional para el diagnóstico correcto en un paciente del cual se sabe que está enfermo (ej.: raspaje de piel en un gato con lesiones cutáneas).
- g) Una prognosis correcta en un paciente del cual se sabe la enfermedad que lo aqueja (ej.: medición del nivel de uremia en un perro con diagnóstico clínico de síndrome urémico).
- h) Una determinación de la existencia de factores de riesgo para una probable futura enfermedad (ej.: medición del colesterol sanguíneo en el hombre como factor de riesgo de enfermedad coronaria).

Bibliografía

- Abdala, A.A., Tarabla, H.D., Bertero, S., Torres, P. (1999).** Vigilancia epidemiológica de Tuberculosis en el Dpto. Castellanos Santa Fe (Argentina). *Rev. Med. Vet.* **80**: 357–360.
- Abdala, A.A. y Tarabla, H.D. (2009).** Detección de rodeos lecheros con hidatidosis a partir de información proveniente de frigorífico. *Rev. FAVE*, **8(2)**: 23–27.
- Abramson, J.H. (1988).** Making sense of data. Oxford Univ. Press, New York, 326 pp.
- Dohoo, I.R., Wright, P.F., Ruckerbauer, G.M., Samagh, B.S., Robertson, F.J. and Forbes, L.B. (1986).** A comparison of five serological tests for bovine brucellosis. *Can. J. Vet. Res.* **50**: 485–493.
- Dohoo, I., Martin, W., Stryhn, H. (2003).** Veterinary Epidemiology Research. Ed. AVC Inc., Charlottetown, Prince Edward Island (Canadá), 706 pp.
- Galen, R. S. and Gambino, S. R. (1975).** Beyond Normality, the Predictive Value and Efficiency of Medical Diagnoses. John Wiley & Sons, New York, 237 pp.
- Knox, K.M.G., Reid, S.W.J., Irwin, T., Murray, M., Gettinby, G. (1998).** Objective interpretation of bovine clinical biochemistry data: application of bayes law to a database model. *Prev. Vet. Med.* **33**: 147–158.
- Last, J. M. (1983).** A Dictionary of Epidemiology. Oxford Univ. Press, New York, 114 pp.
- Marcos, E.R. (1982).** Valores poblacionales de parámetros sanguíneos según el estado de lactancia y época del año. INTA, EERA Rafaela, Pub. Téc. N1 23, 23 pp.
- Michell, A. R. (1992).** Hypocalcaemia: New solutions for old bottlenecks. *Br. Vet. J.* **148**: 271–273.
- Pérez, A.M., Ward, M.P., Torres, P. and Ritacco, V. (2002).** Use of spatial statistics and monitoring data to identify clustering of bovine tuberculosis in Argentina. *Prev. Vet. Med.* **56**: 63–74.
- Real Academia Española (1984).** Diccionario de la Lengua Española. Tomo II. Ed. Espasa-Calpe, S.A., 201 Ed., 1416 pp.
- Smith, R. D. (1991).** Veterinary Clinical Epidemiology. A problem-oriented approach. Butterworth-Heinemann, Reed Pub. Inc. USA, 225 pp.
- Tarabla, H.D. (1995).** Pruebas diagnósticas. Uso e interpretación epidemiológica. En: La Bruce-losis en Sistemas de Producción de Leche, INTA : 1–19.
- Tarabla, H.D. y Abdala, A.A. (1998).** Uso de información proveniente de frigoríficos en la vigilancia epidemiológica de enfermedades notifiables. XII Reu. Cien. Téc. Asoc. Arg. Vet. Lab. Diag., Mar del Plata.
- 16. Tarabla, H. D. y Fernández, F. 1992.** Uso de la prueba VIAA en la vigilancia epidemiológica de la fiebre aftosa. *Bln. Inf. Col. Mjd. Vet. Santa Fe*, **10** : 10–11.

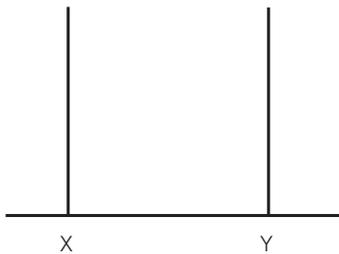
Capítulo 3

Características de las pruebas

1. Factores que afectan la capacidad discriminatoria de la prueba

El principio fundamental de las pruebas diagnósticas se basa en que los individuos que tienen una enfermedad son distintos de los que no la tienen y que las pruebas diagnósticas permiten distinguir a los dos grupos (Figura 1).

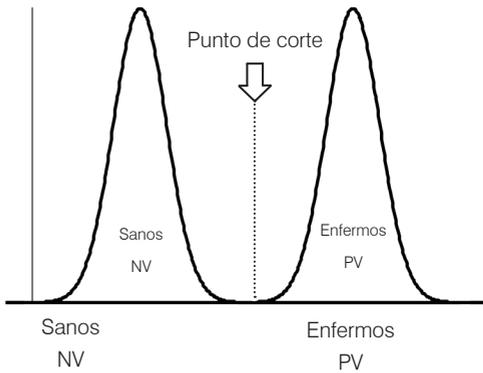
Figura 1. Distribución de valores de la prueba diagnóstica para individuos sanos (X) y enfermos (Y) en una prueba ideal



La prueba diagnóstica ideal podría incluso permitir cierta dispersión en los resultados siempre que cumpla tres premisas (Figura 2) (Duffy, 1991; Riegelman y Hirsch, 1989:256).

- a) Todos los individuos sanos tienen un rango de valores uniforme en los resultados de la prueba diagnóstica.
- b) Todos los individuos enfermos tienen un rango de valores uniforme en los resultados de la prueba diagnóstica, pero diferente de los obtenidos en individuos sanos.
- c) Todos los resultados de la prueba diagnóstica son consistentes con la condición de individuo sano o enfermo.

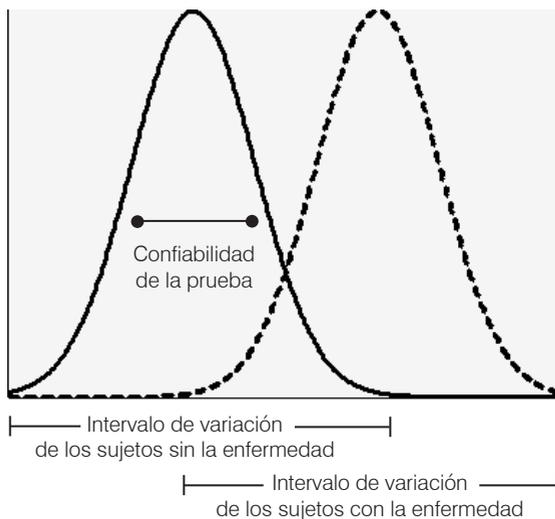
Figura 2. Representación gráfica de una población animal con respecto a su estado de salud y los resultados de una prueba diagnóstica ideal



PV = Positivos verdaderos **NV** = Negativos verdaderos

En un mundo ideal, si es que podemos hablar de ello, la prueba perfecta permitiría diferenciar sin errores enfermedad de salud y el trabajo del clínico se vería limitado a solicitar la prueba adecuada. Sin embargo, el mundo real es más complejo y habitualmente no están presentes las tres condiciones antes señaladas. Las fluctuaciones en los resultados de las pruebas reflejan el efecto combinado de numerosos fenómenos. Existen variaciones en tres factores básicos: la prueba, el grupo de enfermos y el de sanos (Figura 3).

Figura 3. Tipos de variaciones que afectan a las pruebas diagnósticas (adaptado de Riegelman y Hirsch, 1989:256)



Por ello, la interpretación de esos resultados dependerá de la habilidad del clínico para identificar esas fuentes de variación (Knapp y Clinton Miller, 1992:435). La capacidad discriminatoria de una prueba depende de la variabilidad de la misma prueba, de la variabilidad de los individuos sanos y de la variabilidad de los individuos enfermos (Riegelman y Hirsch, 1989:256).

1.1. Variabilidad de la prueba diagnóstica

Los trabajos publicados sobre evaluación de pruebas diagnósticas frecuentemente utilizan términos tales como eficiencia, eficacia, efectividad, utilidad, valor, validez o exactitud. La palabra «diagnóstico(a)» usualmente acompaña a estos términos (ej.: eficacia diagnóstica o valor diagnóstico). Sin embargo, sus significados son, con frecuencia, vagos o variables y la misma palabra parece tener distintos significados para diferentes autores (Yamamoto, 1983). Esto se complica aún más con las traducciones literales o conceptuales de estos términos a diferentes idiomas. A continuación se intentará ordenar los conceptos más utilizados, sin pretender que este ordenamiento sea el único posible. Algunas de las características que se describen a continuación están referidas más al control de calidad discriminatoria de la prueba dentro del laboratorio o entre laboratorios que a la medición cuantitativa del poder discriminatorio. Esta última (ej.: la sensibilidad y especificidad de la prueba) será tratada en el Capítulo 5.

Confiabilidad (ausencia de error aleatorio) y validez (ausencia de error sistemático o sesgo) son elementos independientes de toda prueba. Un proceso diagnóstico puede repetir consistentemente resultados falsos positivos (confiable pero no válido) o puede dar resultados correctos con relativamente alta variabilidad entre sub-muestras (válido pero no confiable). Evaluar una prueba diagnóstica implica confirmar su confiabilidad y su validez, mientras que validarla se aplica más a la medición de su capacidad discriminatoria. Generalmente se refiere a la falta de validez como una falla de la prueba en detectar individuos verdaderamente enfermos (poco sensible) y/o verdaderamente sanos (poco específica), aunque una baja confiabilidad también pueda afectar su sensibilidad y su especificidad (Martin et al., 1987:343). Confiabilidad y precisión están asociadas con la ausencia de error aleatorio, mientras que validez se relaciona con la ausencia de sesgo. Por último, exactitud se relaciona con la ausencia de ambos tipos de errores.

1.1.1. Exactitud

Exactitud de una prueba es la capacidad que tiene de producir resultados cercanos a la verdadera medida del fenómeno anatómico, fisiológico o bioquímico que está sujeto a medición. En la práctica, puede ser estimada por la concordancia observada entre los resultados de la nueva prueba y aquella de referencia. Una prueba para diagnosticar enfermedad que tenga gran exactitud reflejará en forma muy aproximada el verdadero estado de salud del individuo. Para ello se requiere que la prueba sea a la

vez confiable y sin desviaciones sistemáticas a diferir del verdadero valor en una dirección determinada, como se verá más adelante en este capítulo.

1.1.2. Precisión

Algunos autores utilizan este término como sinónimo de confiabilidad, pero diferencia en el grado de detalle de la medición. Una prueba puede reflejar el valor verdadero (ser exacta) pero sin un alto grado de detalle. Por ejemplo, para detectar hipertermia podemos apoyar el dorso de la mano sobre el hocico del animal. Utilizar un termómetro clínico de vidrio con columna de mercurio, ahora en progresivo desuso, con una lectura de 40,2 °C reflejaría con mayor exactitud la alta temperatura corporal. Sin embargo, un termómetro digital podría detectar con mayor precisión una temperatura de 40,234 °C. La falta de precisión causa errores de medición aleatorios.

1.1.3. Validez

Es una expresión del grado en el cual una prueba mide lo que supone debe medir. En ocasiones una prueba es utilizada incorrectamente para reflejar situaciones para las cuales no fueron diseñadas. Por ejemplo, cuando se utilizan los conteos de células somáticas (CCS) individuales para determinar la prevalencia de infecciones intramamarias en vacas lecheras en el comienzo de su lactancia. Esta medición puede ser de ayuda en la evaluación de un programa de control de mastitis dado que refleja la presencia o ausencia de un problema de infecciones intramamarias en el rodeo. Sin embargo, la prueba sólo es válida después del noveno día de lactancia en vaquillonas y después del día 11 en vacas multíparas, debido básicamente al alto número de falsos positivos que se observan en los días anteriores (Dohoo, 1993). La falta de validez provoca errores de medición sistemáticos (ver sesgo).

La precisión y la validez son características independientes de cualquier técnica diagnóstica. Una prueba puede ser precisa (medir siempre el mismo resultado con pequeñas variaciones) pero no válida (producir sólo resultados falsos).

1.1.4. Confiabilidad

Es el grado de estabilidad exhibida cuando la prueba es repetida en condiciones idénticas (Last, 1983:114) y obviamente también afecta la exactitud de la medición. Confiabilidad se refiere al grado en el cual los resultados obtenidos en una medición pueden ser replicados. Una prueba es confiable si los resultados son idénticos o muy similares cada vez que se efectúa. Por lo tanto, una buena prueba produce resultados similares al ser repetida bajo condiciones semejantes. La confiabilidad está influenciada por:

- a) La variabilidad de las condiciones del paciente y del laboratorio.
- b) La variabilidad de interpretación entre observadores.
- c) La variabilidad de interpretación del mismo observador en momentos diferentes.

El primer factor **(a)** incluye las divergencias en el instrumental utilizado y la inestabilidad del objeto sujeto a observación. Los valores obtenidos en la medición de la mayor parte de las características biológicas en los animales varían con el tiempo, incluso durante cortos períodos (ej.: recuento de glóbulos rojos, temperatura corporal, lípidos en sangre). Los resultados de las pruebas diagnósticas pueden incluso variar en función del momento del día en el cual se realizan (ej.: glucosa sanguínea post-prandial o post-ejercicio, o valores obtenidos a partir de muestras de sangre obtenidas en el hogar o en la clínica veterinaria). Por lo tanto, antes de evaluar cualquier resultado se deben considerar las condiciones bajo las cuales se realizó la prueba diagnóstica. Este primer factor puede, a su vez, afectar a los dos siguientes **(b y c)**.

El segundo factor **(b)**, la variabilidad entre observadores, también ha sido llamado concordancia o variabilidad inter-observador. Se refiere a la habilidad de diferentes personas de obtener el mismo resultado para una prueba dada ejecutada sobre la misma muestra. Esta variabilidad aumenta con la subjetividad de la prueba y disminuye con el nivel de entrenamiento del personal (ej.: diagnóstico de un tumor a partir de la observación mediante ecografía).

El tercer factor **(c)**, la variabilidad del mismo observador, también conocida como variabilidad intra-observador, es la habilidad de obtener el mismo resultado en pruebas repetidas sobre la misma muestra. Pruebas como las ecografías, las placas de rayos X o las tomografías pueden ser interpretadas de manera diferente por el mismo observador; comenzando a jugar muchos factores de apreciación subjetiva.

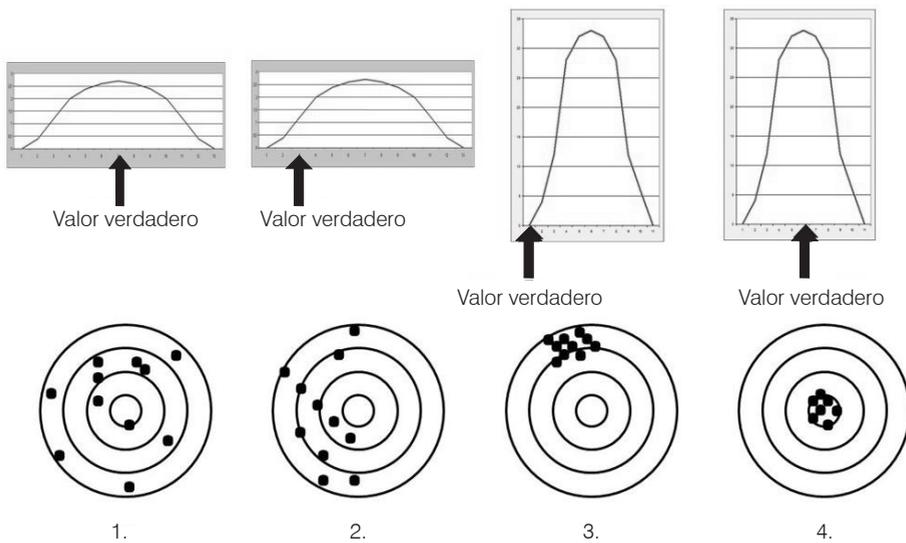
Se debe hacer notar que existen discrepancias entre diversos autores sobre el uso de algunos términos. Por ello, se considera prudente alertar sobre la diversidad de términos utilizados para un mismo fin. Confiabilidad fue también llamada reproducibilidad (Abramson, 1988:326; Riegelman y Hirsch, 1989:256), repetitividad (Abramson, 1988:326; Cannon y Roe, 1982:35), estabilidad (Cronbach, 1951), equivalencia (Cronbach, 1951) y precisión (Martin et al., 1987:343; Palmer y Cavallaro, 1980). En el uso diario, sin embargo, los términos reproducibilidad y repetitividad son utilizados como sinónimos de confiabilidad, aunque se refieren más al hecho de repetir la prueba más que a la calidad de la medición efectuada por la misma. *Repetitividad* es el grado de variación entre los resultados obtenidos para una misma muestra dentro de un mismo laboratorio. Por su parte, *reproducibilidad* es el grado de variación en los resultados de una misma muestra obtenidos por dos o más laboratorios. De cualquier manera, ambos se refieren a la capacidad de la prueba de producir el mismo resultado en cada ocasión que se efectúa la prueba en condiciones similares. Por otra parte, estabilidad y equivalencia fueron utilizadas en la evaluación de test psicológicos como sinónimos de reproductibilidad y concordancia, respectivamente. Con referencia al uso del término precisión, consideramos conveniente conservarlo para medir el grado de error aleatorio.

Por último, no se debe confundir confiabilidad (*reliability*) con confianza (*confidence*). La primera marca la variación de los resultados de un mismo instrumento en manos de diferentes operadores o en momentos distintos, mientras que la segunda está referida a la credibilidad de la información obtenida (Almeida, 1992:112).

1.1.5. Sesgo

Es un error o desviación sistemática del valor verdadero. Un resultado afectado por sesgo influye directamente la exactitud del mismo. En la Figura 4 se representa la relación existente entre el sesgo y la confiabilidad y su efecto en la exactitud de la prueba.

Figura 4. Relaciones entre la validez, el sesgo, la confiabilidad y la exactitud de la prueba



1. Validez alta, sesgo bajo, confiabilidad baja: exactitud baja
2. Validez baja, sesgo alto, confiabilidad baja: exactitud baja
3. Validez baja, sesgo alto, confiabilidad alta: exactitud baja
4. Validez alta, sesgo bajo, confiabilidad alta: exactitud alta

1.2. Variabilidad de los individuos sanos

Si aplicamos una prueba diagnóstica a una población de individuos sanos obtendremos una serie de resultados. Estos seguramente no tendrán un valor idéntico debido a variaciones dentro del mismo individuo y entre individuos. Las primeras dependen de la variación biológica verdadera, los cambios en las condiciones bajo las cuales se realiza la medición o los errores del instrumento de medición. Las segundas, por otra parte, están influenciadas por las variaciones biológicas verdaderas entre indi-

viduos, las diferencias sistemáticas en las condiciones bajo las cuales se realiza la medición y a errores de medición (Knapp y Clinton Miller, 1992:435).

1.2.1. Significados de «normal».

La variación biológica verdadera es la suma de muchos factores desconocidos, cada uno de los cuales contribuye a un pequeño efecto aleatorio que lleva a la obtención de los distintos valores obtenidos en cada individuo (Knapp y Clinton Miller, 1992:435). Esta fluctuación nos hará preguntar: ¿este resultado es normal? La respuesta dependerá del significado que le demos al término «normal». En el Cuadro 1 se presentan algunas de las acepciones que puede tener el mencionado vocablo.

Cuadro 1. Acepciones de «normal»

Significado	Disciplina	Términos preferidos
Curva campana	Estadística	Gaussiana
Más representativo	Ciencias descriptivas	Media, mediana, modo
Más frecuente	Ciencias descriptivas	Habitual
Que se halla en su estado natural	Ciencias naturales	Virgen, natural
Más apto para sobrevivir	Genética, cirugía	Óptimo
Que no daña	Clínica médica	Inocuo
Que sirve de norma o regla	Política, sociología, derecho	Normativo, convencional
Más perfecto	Metafísica, ética, moral	Ideal

Fuentes: Murphy (1972), Real Academia Española (1984).

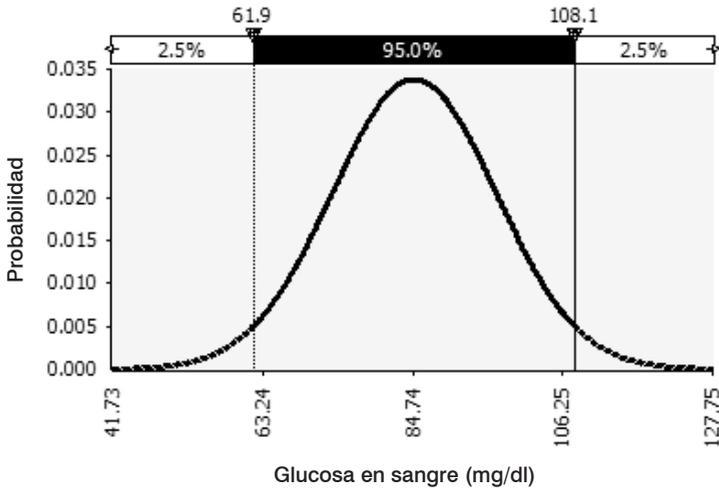
1.2.2. Determinación de normalidad

La selección del criterio de positividad (anormal) y negatividad (normal) para una prueba diagnóstica determinada se puede efectuar utilizando seis metodologías diferentes (Knapp y Clinton Miller, 1992:435).

a) Distribución gaussiana

El intervalo de lo normal es un intento para tratar con la variabilidad que existe entre los animales. Se basa en el supuesto que los resultados de la prueba siguen una curva gaussiana. Para definir el rango normal se calculan los límites superiores e inferiores como la media aritmética ± 2 desvíos estándar. En una población sana, el 95 % de los valores caerán dentro de este intervalo y son clasificados como normales, mientras que el 5 % restante es definido como anormales (Figura 5).

Figura 5. Distribución normal de la concentración de glucosa en sangre en perros



Aunque posee la ventaja de la simplicidad, tiene las siguientes desventajas:

- Los resultados de una prueba diagnóstica frecuentemente tienen distribuciones asimétricas o bimodales, por lo que el rango normal no necesariamente contiene el 95 % de los valores (Knapp y Clinton Miller, 1992:435).
- Si el 5 % de los resultados son siempre clasificados como anormales, todas las enfermedades tendrán la misma frecuencia esperada
- Las pruebas múltiples incrementan la probabilidad de que un paciente caiga fuera del rango normal (ver Capítulo 10.7.1. «Paneles de pruebas»).
- La definición de enfermedad carece de base biológica, dado que el rango es definido en la población sana; un individuo clasificado como anormal no es necesariamente enfermo y uno normal no es necesariamente sano.
- El individuo puede pertenecer a un subgrupo con un rango normal diferente del grupo utilizado para comparar (Greiner et al., 1981).
- Los cambios en los valores obtenidos en la prueba en diferentes momentos pueden ser de significancia patológica aunque se mantengan dentro del rango «normal» (ej.: rápidas fluctuaciones del azúcar en sangre).
- No tiene en cuenta los múltiples objetivos que puede alcanzar la aplicación de una prueba (Greiner et al., 1981).

b) Percentiles

Luego de obtener los valores en un gran número de individuos sanos, se define como anormal el 5 % de los valores más altos, o bien el 2,5 % superior y el 2,5 % inferior de la distribución. El 95 % restante se define como normal. Para los fines prácticos, si el supuesto de distribución gaussiana es verdad, los estimadores no paramétricos (percentiles con sus correspondientes intervalos de confianza) serán tan precisos como

los estimadores paramétricos, aunque requerirán un mayor número de muestras. No obstante, si la distribución no es gaussiana, los estimadores no paramétricos serán más precisos (Reed et al., 1971). Este método es simple y aplicable a todo tipo de distribuciones, pero tiene todas las otras desventajas del método gaussiano enumeradas en el punto anterior.

c) Aceptación cultural

En este caso se define como normal lo que es socialmente aceptado por los patrones culturales de la época. Por ejemplo, las diversas variedades de perros hoy reconocidas como razas caninas se formaron a lo largo del tiempo por efecto de una conjunción de factores geográficos, climáticos y sociales. El Setter Irlandés en sus inicios era blanco con manchas rojizas. No obstante, el deseo de obtener un ideal de belleza hizo que esas manchas se fueran extendiendo transformando el color del animal en un rojizo cálido con manchas blancas inexistentes o reducidas a límites muy pequeños. Por la constante selección a la que ha sido sometido a lo largo del tiempo, un Setter Irlandés original no sería actualmente aceptado en ninguna exposición canina. Por otra parte, un ejemplo típico en humanos es el rango de peso normal entre las mujeres que está definido por el consenso popular de delgadez. Este método no presenta ventaja alguna y confunde medicina con otras disciplinas.

d) Terapéutico

Define como anormal los valores por encima o debajo del cual el individuo requiere tratamiento. Si bien clasifica como anormal a los individuos que realmente van a ser tratados, los límites de normalidad varían de acuerdo con el conocimiento de las restricciones o beneficios del tratamiento. Por ejemplo, el tratamiento simultáneo durante la lactancia de las infecciones intramamarias subclínicas existentes en un rodeo lechero generó inicialmente buenas expectativas como herramienta de lucha contra la mastitis bovina. Sin embargo, nuevos trabajos restringieron su utilidad a las infecciones causadas por *Streptococcus agalactiae*. Si bien este criterio primero clasificó como anormal a todos los animales con infecciones intramamarias de disímil etiología, el conocimiento de las limitantes del tratamiento acotó el mismo a sólo un agente causal (Tarabla, 1990). Por el contrario, en el caso de la hipertensión arterial en humanos la proporción de tratados fue en aumento, dado que los límites para ser considerado hipertenso han sido disminuidos progresivamente a medida que nuevos hallazgos demostraron la conveniencia de tratar valores cada vez más bajos.

e) Factores de riesgo

En este procedimiento se asume que la asociación existente entre el factor de riesgo y la enfermedad puede predecir la existencia de la misma. «Normalidad» en estos casos se refiere a los valores de la prueba diagnóstica que no adicionan un mayor riesgo de morbilidad o mortalidad. Este método tiene las ventajas que los factores de

riesgo son generalmente fáciles de medir y que la identificación con una enfermedad específica facilita las medidas de prevención y control. Sin embargo, la probabilidad de que un individuo enfrentado al factor de riesgo esté realmente enfermo puede ser muy baja. Por ejemplo, la presencia de áreas encharcadas es un factor de riesgo conocido para la ocurrencia de leptospirosis bovina. No obstante, la mera presencia de charcos en un campo no tiene valor diagnóstico de la enfermedad en un individuo. Igualmente, el hábito de fumar aumenta significativamente el riesgo de enfermedades pulmonares y circulatorias en humanos pero, sin embargo, el sólo hecho que el paciente sea fumador tiene escaso valor para predecir la presencia de enfermedad.

f) Diagnóstico

Este método, también llamado del valor predictivo, define la probabilidad de enfermedad a través de la comparación con una prueba patrón y será discutido en profundidad en los próximos capítulos. A partir de ellos se define el criterio de normal (negativo a la prueba diagnóstica) y anormal (positivo a la misma prueba). Sus ventajas son que los resultados de la prueba pueden tener cualquier tipo de distribución, que se tienen en cuenta la distribución de los resultados de la prueba tanto en la población enferma como en la sana y que la frecuencia de enfermedad es una estimación realista del parámetro verdadero. Como desventajas se pueden señalar la necesidad de conocer la prevalencia de la enfermedad y las frecuencias de resultados positivos entre los enfermos y de negativos entre los sanos.

1.2.3. Normal y anormal. Ejemplos

Tomemos como ejemplo el dosaje de cobre (Cu) sanguíneo en un supuesto laboratorio ubicado en la zona de cría bovina del norte de la provincia de Santa Fe, Argentina. Si tomamos como normales los valores que se encuentran dentro de ± 2 desvíos estándar de la media en la distribución de frecuencias de los valores encontrados en el área de influencia del laboratorio, veremos que tiene sus riesgos. Estos valores no nos serán de gran utilidad para el diagnóstico de hipocupremia en el ganado bovino, dado que el área norte santafesina es conocida por el problema de hipocuprosis. A similares conclusiones podríamos arribar, por ejemplo, con la medición de la presión arterial en poblaciones humanas. Normal en una población urbana occidental tendrá indudablemente valores más altos que lo normal en una población indígena con menor exposición a los factores de riesgo conocidos para la hipertensión arterial (alimentación sazónada con cloruro de sodio, estrés, obesidad, sedentarismo). Evidentemente, en estos ejemplos extremos, la presentación de hipocuprosis en bovinos o de accidentes hipertensivos en humanos, es independiente de los valores estadísticamente normales en los individuos sanos residentes en el área geográfica. Esto último está relacionado simplemente con la elección del grupo de referencia que toma el investigador y en su criterio para seleccionar el grupo de individuos que considera no están enfermos. Por

otra parte, la población de referencia usada para calcular los límites no está necesariamente libre de la enfermedad (Greiner et al., 1981). Más aún, de acuerdo con el ya visto Cuadro 1, normal para los valores de cupremia podría tomar cualquiera de los siguientes significados:

- a) Una distribución de frecuencias de los valores de cupremia en forma de campana.
- b) El valor más representativo de cupremia y definido por la media, la mediana o el modo.
- c) El valor comúnmente encontrado y definido por un rango.
- d) El valor de cupremia en su estado original.
- e) El valor de cupremia más apropiado para reproducirse o sobrevivir.
- f) Valores de cupremia que no causen daño.
- g) El consenso de una comisión o comité. Valores de cupremia “aprobados”.
- h) El valor de cupremia ideal.

El laboratorista usualmente está interesado en los significados **a)** o **c)**, usualmente utilizados como «Rango Normal de Laboratorio». Sin embargo, el veterinario de campo que atiende el caso clínico y espera la confirmación de su diagnóstico presuntivo seguramente estará más interesado por saber si el animal sobrevivirá o podrá gestar (significado **e)**), o bien en si ese nivel de Cu en sangre resultará perjudicial para el individuo (significado **f)**) o incluso si ese es el valor ideal de la cupremia (significado **h)**).

Desde el punto de vista de los significados **a)** y **c)** representan un concepto que mide y cuantifica el intervalo de valores que existen en individuos considerados sanos. Este rango normal es un mal necesario y está basado en la suposición de que el individuo bajo estudio debe ser similar a otros individuos. No es un concepto diagnóstico, sino descriptivo, es decir, no diagnostica la enfermedad sino que describe a los individuos sanos. Los valores que se encuentran fuera de este rango pueden ser el resultado de cambios causados por la enfermedad, pero también de variaciones aleatorias, de cambios fisiológicos no asociados con la enfermedad, o de cambios patológicos secundarios a la misma (Riegelman y Hirsch, 1989:256). Por ello, un resultado normal no siempre significa sano, un resultado anormal no siempre significa enfermo, y la presencia de cambios, aunque se encuentren dentro del rango normal, puede ser indicativa de enfermedad (Duffy, 1991). Este concepto difiere del clásico del laboratorio clínico que define la población normal exclusivamente por el promedio y el desvío estándar, postulando que si el resultado obtenido queda fuera de tres desviaciones estándar «es evidente que este individuo no pertenece a la población normal» (Cope-land, 1951, 1287). En realidad, en las distribuciones de los resultados positivos y negativos existe un área donde se superponen las poblaciones sanas y enfermas, lo que será discutido en el Capítulo 4.

1.3. Variabilidad de los individuos enfermos

La obtención de un resultado anormal puede deberse a cambios patológicos en el individuo. Sin embargo, si aplicamos una prueba diagnóstica a una población de individuos enfermos tampoco obtendremos resultados idénticos, sino una serie de resultados que pueden distribuirse en forma no gaussiana y que variarán de acuerdo con la habilidad de cada individuo para producir una respuesta medible a través de una prueba diagnóstica. Esta respuesta (ya sea en forma de anticuerpos, reacción tisular u otro tipo de modificaciones) dependerá de factores tales como la estructura genética del hospedador (especie, raza, defectos o resistencia hereditarias), su edad y estado fisiológico, su madurez inmunológica, su estado inmunitario previo, la presencia de factores concurrentes tales como estrés ambiental, estado nutricional, medicación o enfermedades y la vía de entrada, la patogenicidad y la dosis del agente a la que fue expuesto el individuo.

2. Características deseables en una prueba diagnóstica

2.1. Alta capacidad discriminatoria

Una prueba debe tener la capacidad de diferenciar individuos sanos y enfermos. Para que sea considerada una verdadera prueba, la probabilidad de clasificar correctamente un individuo como sano o enfermo debe ser mayor que la probabilidad dada simplemente por el azar. La medición de la capacidad discriminatoria de la prueba será discutida en el Capítulo 5.

2.2. Economía

Este aspecto debe ser visto desde varios puntos de vista. Uno de ellos es el uso de pequeñas cantidades de la muestra de material biológico (sangre, suero, leche, etc.), que permite realizar varias pruebas sobre la misma muestra. La toma de muestras en relevamientos requiere generalmente la disponibilidad de personal entrenado, materiales especiales, movilidad, tiempo y recursos económicos, por lo que es de suma importancia la utilización eficiente del material biológico objeto del muestreo. Paralelamente, también es importante que la prueba necesite la menor cantidad posible de antígeno. En ocasiones, la obtención de antígeno requiere de operaciones complicadas o costosas, por ello es necesario economizar dicho material una vez obtenido. Por ejemplo, la obtención de antígeno para diagnóstico de *Anaplasma marginale* puede implicar el trabajo de un equipo entrenado en pasajes rápidos de sangre infectada en terneros esplenectomizados (Anziani et al., 1980), en otros casos los antígenos son producidos por un sólo laboratorio (un ejemplo era el antígeno empleado para prueba VIA en fiebre aftosa), o son muy caros para ser utilizados masivamente en poblaciones animales (ej.: determinación del perfil metabólico en vacas lecheras). Por último, es importante que la prueba sea económica en otros términos, no sólo de materiales de laboratorio, sino también en la dedicación de recursos humanos (que ocupe poco

personal), del tiempo disponible (que sea rápida) y en recursos económicos (que sea barata). Estos aspectos adquieren particular importancia en casos de utilización masiva de las pruebas para la estimación de la prevalencia de enfermedades.

2.3. Disponibilidad de materiales

Obvio es decirlo: los materiales necesarios para la prueba deberían ser de fácil obtención y estar disponibles para su utilización. Como se desprende del punto anterior, esto no siempre es factible. Por ejemplo, la prueba del gamma interferón para el diagnóstico de tuberculosis bovina requiere de un antígeno que no está fácilmente disponible en todos los mercados.

2.4. Simplicidad

La facilidad con que una prueba es aprendida y aplicada es una consideración de importancia en lugares donde no hay personal técnico entrenado disponible y en trabajos de relevamientos donde es necesario analizar un gran número de muestras. Antes de elegir, siempre es útil recordar que una prueba diagnóstica «cuanto más compleja ... más variables ... más errores» (Tizard, 1982; Yamamoto, 1983).

2.5. Objetividad

La objetividad para la lectura del resultado de una prueba influye directamente en la variabilidad de los resultados obtenidos por diferentes observadores. Por ejemplo, la interpretación de los resultados para detectar individuos con anticuerpos contra *Babesia bovis* será más objetiva si se utiliza la prueba de ELISA que si se usa la prueba de inmunofluorescencia indirecta (Echaide et al., 1995). De igual manera, en el diagnóstico de mastitis subclínicas por métodos indirectos, será dable esperar una menor variabilidad inter-observador si se utiliza un contador automático de células somáticas que si se usa una prueba de aglutinación como la de California. En la primera el resultado es simplemente un número, mientras que en la segunda se requiere de una evaluación por parte del operador del grado de aglutinación observable en la mezcla de la leche y el reactivo.

2.6. Seguridad

La utilización de antígenos vivos en manos inexpertas puede resultar en infecciones serias. Por otra parte, el uso de antígenos vivos en laboratorios sin las normas de seguridad requeridas puede representar el escape de material infeccioso a poblaciones susceptibles cercanas. Por ejemplo, la prueba patrón en leptospirosis, la microaglutinación de Martin y Petit requiere de un cepario que sólo puede ser mantenido correctamente en un laboratorio con estrictas normas de seguridad.

Bibliografía

- Abramson, J.H. (1988).** Making sense of data. Oxford Univ. Press, New York, 326 pp.
- Almeida, N. de. (1992).** Epidemiología sin números. OPS, OMS, Serie PALTEX N1 28, 21 Ed., Buenos Aires, 112 pp.
- Anziani, O.S., Tarabla, H.D., Ford, C.A. y Castro, J.C. (1980).** Pasajes rápidos de *Anaplasma marginale* en terneros esplenectomizados para la obtención de antígeno. III Congr. Arg. Cien. Vet. y III Simposio Nac. Cien. Tec. de Carnes, Buenos Aires, p. 53.
- Cannon, R.M. and Roe, R.T. (1982).** Livestock Disease Surveys: A Field Manual for Veterinarians. Aust. Bureau of Anim. Hlth., Dpt. Primary Ind., Aust. Gov. Pub. Serv., Canberra, 35 pp.
- Copeland, B.E. (1951).** Recursos estadísticos en el laboratorio clínico. En: Davidsohn, I. y Henry, J.B. 1974. Diagnóstico Clínico por el Laboratorio. Salvat Editores S.A, 51 Ed., 31 Reimp., 1287 pp.
- Cronbach, L.J. (1951).** Coefficient Alpha and the internal structure of tests. Psychometrika, 16: 297–334.
- Cunningham, J.G. (1997).** *Textbook of veterinary physiology*. 2nd ed. W.B. Saunders, Philadelphia, EE.UU.
- Dohoo, I.R. (1993).** An evaluation of the validity of individual cow somatic cell counts from cows in early lactation. Prev. Vet. Med. 16: 103–110.
- Duffy, S.J. (1991).** Evaluación de pruebas diagnósticas. Mem. VII Reu. An. Asoc. Arg. Vet. Lab. Diag., Buenos Aires, pp. 6–13.
- Echaide, S.T. de, Echaide, I.E., Gaido, A.B., Mangold, A.J. Lugaresi, C.I., Vanzini, V.R. and Guglielmono, A.A. (1995).** Evaluation of an enzyme-linked immunosorbent assay kit to detect *Babesia bovis* antibodies in cattle. Prev. Vet. Med. 24: 277–283.
- Greiner, P.E., Mayewski, R.J., Mushlin, A.I., Greenland, P. (1981).** Selection and interpretation of diagnostic test and procedures. Ann. Int. Med. 94: 553–600.
- Knapp, R.G. and Clinton Miller, M. (1992).** Clinical Epidemiology and Biostatistics. Natl. Med. Series, Williams & Wilkins, Harwal Publ. Co., Pennsylvania, 435 pp.
- Last, J.M. (1983).** A Dictionary of Epidemiology. Oxford Univ. Press, New York, 114 pp.
- Martin, S.W., Meek, A.H. and Willeberg, P. (1987).** Veterinary Epidemiology. Principles and Methods. Iowa St. Univ. Press, Ames, Iowa, 343 pp.
- Murphy, E.A. (1972).** The normal, and the perils of the sylleptic argument. Perspect. Biol. Med. 15: 566–582.
- Palmer, D.F., and Cavallaro, J.J. (1980).** Some concepts of quality control in immunoserology. In: Rose, N.R. and Friedman H. (eds.). 1980. Manual of Clinical Immunology. 2nd. Ed., Am. Soc. for Microbiol., Washington D.C., pp. 1078–1082.
- Real Academia Española (1984).** Diccionario de la Lengua Española. Tomo II. Ed. Espasa-Calpe, SA, 201 Ed., 1416 pp.
- Reed, A.H., Henry, R.J. and Mason, W.B. (1971).** Influence of statistical method used on the resulting estimate of normal range. Clinical Chemistry. 17: 275–284.
- Riegelman, R.K. and Hirsch, R.P. (1989).** Cómo estudiar un estudio y probar una prueba: Lectura crítica de la literatura médica. Little, Brown & Co., 21 Ed., OPS Pub. Cien. 532, 256 pp.
- Tarabla, H.D. (1990).** Tratamiento simultáneo durante la lactancia de infecciones intramamarias subclínicas (Terapia «blitz»). Therios. 17: 142–149.
- Tizard, I. (1982).** Serologic assays. J. Am. Vet. Med. Ass. 181: 1162–1165.
- Yamamoto, R. (1983).** Factors affecting sensitivity and specificity of serologic tests: Some practical considerations and general thoughts. Univ. California Davis, EPM 216 Notes, 5 pp.

Capítulo 4

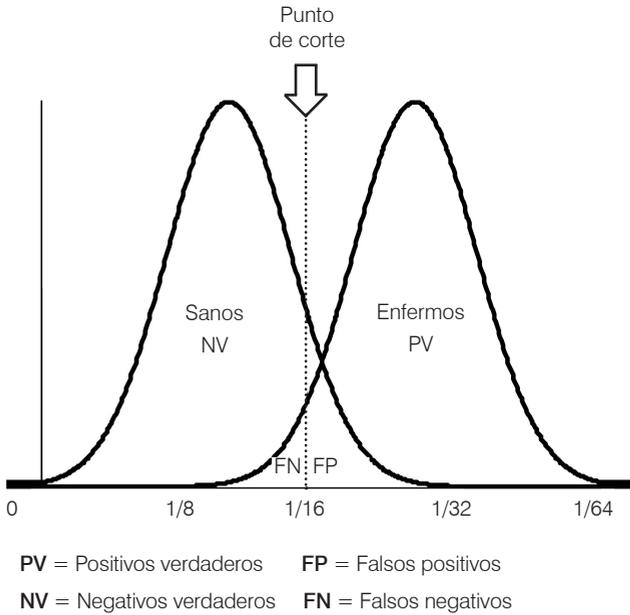
Relación entre el resultado de la prueba y el estado de salud de los individuos

Los resultados de una prueba pueden ser expresados en forma cualitativa («positivos» o «negativos») o por medio de algún tipo de gradiente numérico a lo largo de una escala ordinal o continua. Estos últimos pueden ser expresados en forma dicotómica por medio de la selección de un valor crítico o de corte en la escala por encima del cual los resultados obtenidos se consideran positivos y por debajo del mismo se clasifican como negativos. En algunos casos se incluye una categoría intermedia («sospechosos») que no será considerada en la discusión por razones de simplicidad, aunque las conclusiones finales sean similares. Como se puntualizó en el capítulo anterior, una prueba diagnóstica ideal debería dar resultados completamente diferentes en poblaciones enfermas y sanas. En este caso, todos los individuos enfermos estarían correctamente clasificados por la prueba diagnóstica como positivos (Positivos verdaderos, PV) y todos los sanos estarían correctamente clasificados como negativos (Negativos verdaderos, NV) tal como observamos en la Figura 2 del capítulo anterior.

De esta forma, los resultados de la prueba diagnóstica deberían correlacionarse con el estatus de enfermedad del animal analizado. No obstante, en la vida real no necesariamente existe una completa concordancia. Tomando como ejemplo una prueba que detecte anticuerpos, en una población libre de la enfermedad la mayoría de los animales no tendrán anticuerpos detectables. Sin embargo, algunos de estos animales sanos pueden tener anticuerpos provocados probablemente por reacciones cruzadas con otros agentes. Igualmente en una población enferma, el nivel medio de anticuerpos puede ser relativamente alto, pero inevitablemente algunos animales verdaderamente enfermos tendrán un nivel de anticuerpos bajo (Tizard, 1982). Por ello, si bien la mayoría de los animales donde se realizó la prueba van a estar en alguno de los dos lados de ese valor arbitrario de división llamado punto de corte, siempre existe algún grado de superposición entre las poblaciones sanas y enfermas. En este caso

hay individuos enfermos que dan negativos a la prueba diagnóstica (Falsos negativos, FN) e individuos sanos que dan positivos a la prueba (Falsos positivos, FP) (Figura 1).

Figura 1. Representación gráfica de una población animal con respecto a su estado de salud y los resultados de una prueba diagnóstica



En el ejemplo hipotético de la Figura 1 se seleccionó un valor crítico de 1/16 por encima del cual los animales son clasificados como positivos, incluida la mayoría de los animales enfermos (positivos verdaderos) y una pequeña proporción de animales sanos (falsos positivos). De igual manera por debajo de la dilución 1/16 se incluyen tanto animales sanos (negativos verdaderos) como enfermos (falsos negativos). Una buena prueba es aquella en la cual las distribuciones de frecuencias de los positivos y negativos a la prueba tienen la menor área posible de superposición (Astudillo y Kantor, 1981). Esta relación entre los resultados dicotómicos de una prueba y el estado natural de la población con referencia a una enfermedad también puede ser ilustrada con una tabla de doble entrada en la cual cada celda representa porciones diferentes de las superficies bajo las curvas de animales sanos y enfermos (Cuadro 1).

Cuadro 1. División de una población animal con respecto a su estado de salud y los resultados dicotómicos de una prueba diagnóstica

Condición	Enfermos	Sanos	Total
Test positivos	Enfermos positivos	Sanos positivos	Aparentemente enfermos
Test negativos	Enfermos negativos	Sanos negativos	Aparentemente sanos
Total	Realmente enfermos	Realmente sanos	Población

Esta tabla de doble entrada, simplificada en el Cuadro 2, será de ahora en adelante utilizada como base para la presentación de los conceptos de sensibilidad, especificidad, prevalencia real, prevalencia aparente, valores predictivos y eficiencia.

Cuadro 2. Clasificación de una población animal con respecto a su estado de salud y los resultados de una prueba diagnóstica

Condición	Enfermos	Sanos	Total
Test positivos	PV	FP	PV + FP
Test negativos	FN	NV	NV + FN
Total	PV + FN	NV + FP	PV + FP + NV + FN = n

PV = Positivos verdaderos, **FP** = Falsos positivos,

NV = Negativos verdaderos, **FN** = Falsos negativos, **n** = total de individuos

La sensibilidad y especificidad, que serán tratadas en el Capítulo 5, son características propias de la prueba consideradas como valores fijos por razones de simplicidad. En realidad, constituyen valores promedios referentes al poder discriminatorio de una prueba dada. Por otra parte, la prevalencia real de la enfermedad y la proporción de animales sanos son las probabilidades de tener o no tener la enfermedad y son *a priori* e independientes de la realización de la prueba (ver Capítulo 7). Por último, la prevalencia aparente de la enfermedad, los valores predictivos y la eficiencia de la prueba son probabilidades *a posteriori* de la ejecución de la prueba y serán desarrolladas en los Capítulos 7, 8 y 9. Se debe tener en cuenta que la prevalencia de reaccionantes positivos a una prueba diagnóstica hallada en un estudio también depende de la selección de los individuos incluidos en la muestra y por consiguiente su valor presenta una variación muestral (Marchevsky, 1974).

Bibliografía

Astudillo, V.M. y Kantor, I.N. (1981). El problema de la validez de una prueba diagnóstica para uso masivo como procedimiento estadístico de clasificación. *Bltn. Centr. Panam. Fiebre Aftosa.* **43-44:** 37-43.

Marchevsky, N. (1974). Errores en las estimaciones de prevalencia en estudios de población: Un método para calcular la prevalencia real. *Zoonosis.* **16:** 81-109.

Tizard, I. (1982). Serologic assays. *J. Am. Vet. Med. Ass.* **181:** 1162-1165.

Capítulo 5

Sensibilidad y especificidad

1. Sensibilidad

Sensibilidad es la probabilidad condicional de un resultado positivo en un individuo con la enfermedad $[P(T+ | E+)]$ (Yerushalmy, 1947). En otras palabras, es la habilidad de una prueba para detectar un animal enfermo (*ie.* ser positiva) en una población enferma (Cuadro 1). Cuanto más grande es esta probabilidad, más sensible será la prueba.

Cuadro 1. Sensibilidad de la prueba

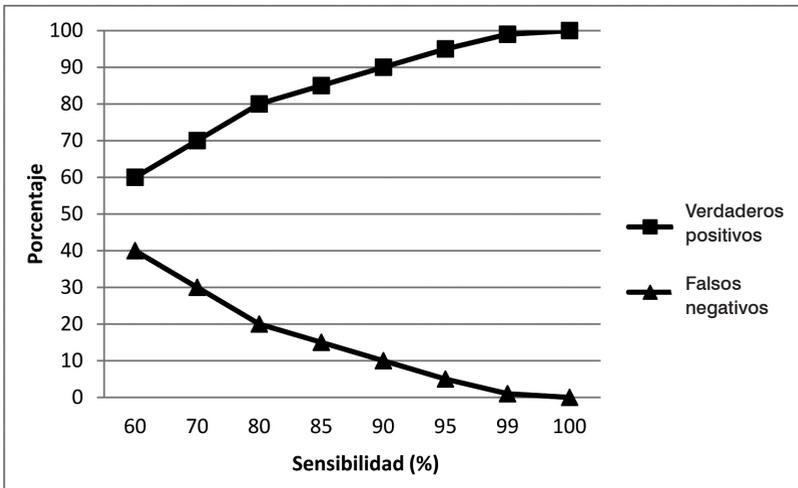
Condición	Enfermos	Sanos	Total
Test positivos	PV	FP	PV + FP
Test negativos	FN	NV	NV + FN
Total	PV + FN	NV + FP	PV + FP + NV + FN = n

$$\text{Sensibilidad} = \text{Sens.} = \frac{PV}{PV + FN} \times 100$$

Sensibilidad mide la capacidad discriminadora de la prueba en una población enferma. Una buena prueba diagnóstica da resultados positivos en individuos enfermos. Nótese que esta definición difiere de la utilizada usualmente en inmunología, toxicología y bioquímica, donde sensibilidad es la capacidad de una prueba de detectar pequeñas cantidades de antígenos, anticuerpos, toxinas, drogas, metabolitos u otros elementos. Ésta es la llamada sensibilidad analítica y se refiere al límite de detección (ej.: mínima concentración de un analito que debe estar presente en la muestra para que la prueba de un resultado positivo) (Houe et al., 2003:340) y si bien

debe reconocerse su existencia y utilidad, la misma no será considerada en el resto del libro ya que es un parámetro que está incluido en la sensibilidad de la prueba diagnóstica. En epidemiología, sensibilidad es la incidencia de resultados verdaderamente positivos obtenidos cuando la prueba es aplicada a individuos con la característica buscada. Una prueba inmunológicamente sensible y específica puede no serlo desde el punto de vista epidemiológico. A medida que aumenta la sensibilidad, aumentan los diagnósticos positivos verdaderos y disminuyen los falsos negativos (Figura 1).

Figura 1. Relación entre sensibilidad, verdaderos positivos y falsos negativos



Mencionaremos diferentes causas que generan una pérdida en la sensibilidad de una prueba diagnóstica:

- a) Tolerancia natural o inducida por parte de algunos animales a las infecciones (ej.: exposición al virus de la Diarrea Viral Bovina al principio de la gestación);
- b) Tiempo: el agente o la respuesta inmune no se logró establecer;
- c) Pobre crecimiento de agentes patógenos microbianos en los medios de cultivo;
- d) Inhibidores no específicos (ej.: suero contaminado o hemolizado);
- e) Anticuerpos bloqueantes (ej.: demasiada IgG1 bloqueando la IgG2);
- f) Inmunosupresión;
- g) Anticuerpos maternos que eliminan el agente infeccioso;
- h) El agente infeccioso se elimina de manera intermitente;
- i) Errores de laboratorio;
- j) Baja sensibilidad analítica.

2. Especificidad

Especificidad es la probabilidad condicional de un resultado negativo en un individuo sin la enfermedad ($P(T-|E-)$). Por ende, es la habilidad de una prueba de detectar un animal sano (ej.: ser negativa) en una población sana (Cuadro 2). Cuanto más grande es esta probabilidad, más específica será la prueba.

Cuadro 2. Especificidad de una prueba

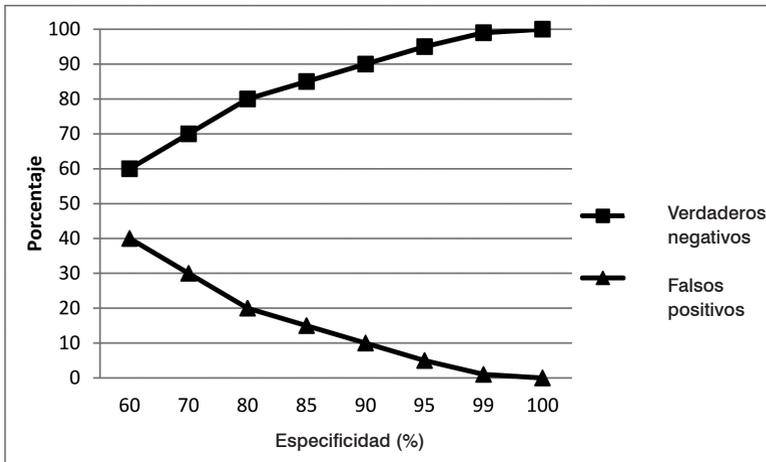
Condición	Enfermos	Sanos	Total
Test positivos	PV	FP	PV + FP
Test negativos	FN	NV	NV + FN
Total	PV + FN	NV + FP	PV + FP + NV + FN = n

$$\text{Especificidad} = \text{Esp.} = \frac{NV}{NV + FP} \times 100$$

No debe confundirse la especificidad epidemiológica tal como se expresó anteriormente con la especificidad analítica, la cual se refiere a la habilidad de una prueba para identificar correctamente o distinguir entre analitos similares (ej.: reacciones cruzadas con anticuerpos similares a los que uno desea detectar) (Houe et al., 2003:340).

Con el término especificidad se denomina la incidencia de resultados verdaderamente negativos obtenidos cuando la prueba es aplicada a individuos sin la característica buscada. Un buen test da resultados negativos en individuos sanos. A medida que aumenta la especificidad, aumentan los diagnósticos negativos verdaderos y disminuyen los falsos positivos (Figura 2).

Figura 2. Relación entre especificidad, verdaderos negativos y falsos positivos



Mencionaremos diferentes causas que generan una pérdida en la especificidad de una prueba diagnóstica:

- a) Relación antigénica entre diferentes agentes infecciosos;
- b) Reacciones inmunes cruzadas;
- c) Reacciones no específicas (ej.: inhibidores de la hemaglutinación);
- d) Reacciones inmunes causadas por la vacunación;
- e) Adquisición de inmunidad pasiva por calostro;
- f) Exposiciones previas;
- g) Errores de laboratorio.

Se puede decir que no existe prueba diagnóstica que sea a la vez 100 % sensible y 100 % específica. Ni siquiera las pruebas utilizadas para diagnosticar muerte en humanos son a la vez 100 % sensibles y 100 % específicas (Galen y Gambino, 1975:237) al ser interpretadas por separado y habiendo sido aplicadas una única vez. En estos casos sensibilidad es la habilidad del test de detectar individuos positivos (muertos) en una población de muertos y especificidad la habilidad del test de detectar individuos negativos (vivos) en una población de vivos. Veamos primero el ejemplo del electroencefalograma (EEG) chato. Todos los individuos sin vida muestran EEG chato; ergo, todos los muertos son positivos a la prueba y, por lo tanto, el test es 100 % sensible. Ahora bien, ¿todos los vivos presentan ausencia de EEG chato? o sea ¿todos los vivos son negativos a la prueba? Seguramente no, dado que no todos los individuos en coma profundo terminan muriéndose; existen por lo tanto algunos que habiendo dado un EEG chato en realidad estaban vivos. Por lo tanto, no todos los vivos son negativos a la prueba o dicho en otras palabras el test no es 100 % específico. Tomemos ahora el ejemplo del *rigor mortis*. Podemos afirmar sin temor a equivocarnos que todos los individuos con vida muestran ausencia de *rigor mortis*, ergo, que todos los vivos son negativos a la prueba y, por lo tanto, que el test es 100 % específico. Sin embargo, ¿todos los muertos presentan *rigor mortis*? Evidentemente no, dado que la presencia de este signo en un individuo muerto depende de numerosos factores tales como el tiempo transcurrido desde la muerte, la temperatura ambiente, etc. Por lo tanto, no todos los muertos son positivos a la prueba o, dicho en otras palabras, el test no es 100 % sensible. Por ello, para detectar muerte en humanos se repiten varias pruebas para maximizar así su valor predictivo (ver Capítulo 10). Esto lleva a obtener un 100 % de especificidad, eliminando la posibilidad de diagnósticos de muerte en la población viva.

Un aspecto muy importante es que la sensibilidad y la especificidad son características intrínsecas de una prueba que deben ser idénticas, ya sea que se aplique a un grupo de animales en los cuales la enfermedad es rara o a un grupo de individuos en los que es frecuente; es decir, la capacidad de discriminación diagnóstica de la prueba debe ser la misma sea cual fuere la probabilidad de enfermedad antes de realizar la prueba.

Por otro lado, hay que tener en cuenta que la sensibilidad y la especificidad solamente indican la proporción o porcentaje de los que han sido correctamente clasificados como sanos o como enfermos. Estas medidas no predicen el número real de individuos que serán clasificados correctamente, cifra que dependerá de la frecuencia de la enfermedad en el grupo al que se le aplique la prueba.

3. Falsos positivos

Son aquellos individuos que, habiendo dado positivos a la prueba diagnóstica, están verdaderamente sanos. La proporción de FP es igual a $1 - \text{Esp.}$ y es análogo al error de Tipo I (probabilidad de α) en inferencias estadísticas, es decir la probabilidad de rechazar la hipótesis nula cuando ésta es correcta (Snedecor y Cochran, 1980:507). En oportunidades se pueden encontrar referencias llamando a los animales falsos positivos como animales «sin lesiones aparentes» o «reactores sanos», haciendo referencia a su negatividad a la prueba patrón, en este caso la ausencia de lesiones patológicas (Seiler, 1979). Si la proporción de falsos positivos se constituye en un problema, se podrá considerar la utilización de pruebas más específicas, la destrucción de las respuestas inmunológicas tempranas presentes en la muestra (ej.: eliminación de inmunoglobulina G por medio del rivanol, la acidez o el calor en suero sanguíneo) (Yamamoto, 1983) o cambios en la concentración del antígeno (Miranda et al., 2002).

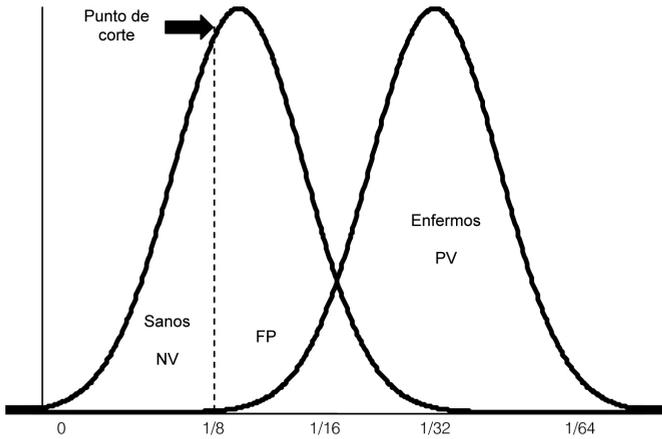
4. Falsos negativos

Son aquellos individuos que habiendo dado negativos a la prueba diagnóstica están verdaderamente enfermos. La proporción de FN es igual a $1 - \text{Sens.}$ y es análogo al error de Tipo II (probabilidad de β) en inferencias estadísticas, es decir la probabilidad de rechazar la hipótesis alternativa cuando ésta es correcta (Seiler, 1979). Si se desea minimizar los falsos negativos se podrá utilizar una prueba más sensible o pruebas que detecten infección temprana (Yamamoto, 1983).

5. Relación entre sensibilidad y especificidad

Definida el área de superposición de las curvas de resultados positivos y negativos en las poblaciones de individuos enfermos y sanos, todo incremento de la sensibilidad de la prueba se hará en detrimento de la especificidad y viceversa (Adler y Wiggins, 1973). En el ejemplo de la Figura 1 del capítulo anterior se había elegido un valor crítico de 1/16. Ahora bien, ¿qué sucede si cambiamos el valor crítico? En la Figura 3 se seleccionó el valor crítico en la dilución 1/8.

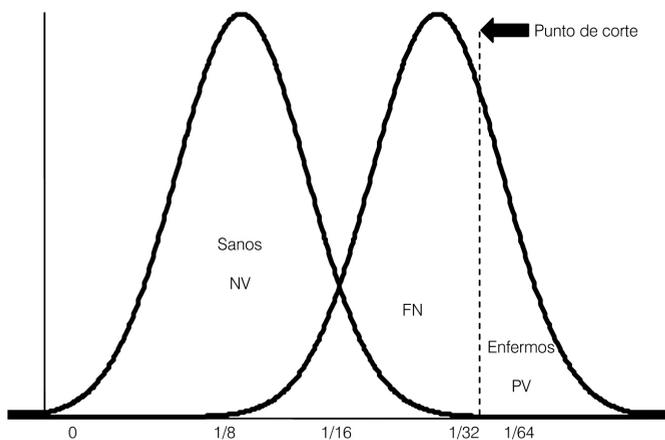
Figura 3. Cambio del valor crítico. Aumento de la sensibilidad y disminución de la especificidad de una prueba diagnóstica



PV = Positivos verdaderos FP = Falsos positivos
 NV = Negativos verdaderos FN = Falsos negativos

En este caso, todos los animales enfermos serán correctamente clasificados como positivos (aumentó la sensibilidad al 100 %). No obstante, habrá un mayor número de animales sanos incorrectamente clasificados como positivos (disminuyó la especificidad). Por otra parte, en la Figura 4 se fijó un valor de corte de 1/64.

Figura 4. Cambio del valor crítico. Aumento de la especificidad y disminución de la sensibilidad de una prueba diagnóstica



PV = Positivos verdaderos FP = Falsos positivos
 NV = Negativos verdaderos FN = Falsos negativos

Todos los animales sanos serán correctamente clasificados como negativos (aumentó la especificidad al 100 %), pero la mayoría de los animales enfermos están incorrectamente clasificados como negativos (disminuyó la sensibilidad).

La importancia de la selección de un nivel de corte bajo o alto dependerá de la importancia que le demos a los falsos positivos y a los falsos negativos. Los falsos positivos están asociados con costos (emocionales y financieros). Los resultados falsos negativos tienen una importante implicancia sanitaria y de posibilidades de curación por no tratar la enfermedad en sus estadios iniciales. De manera general, las pruebas diagnósticas con alta sensibilidad son útiles en los siguientes casos:

- a) En fases tempranas de la patogénesis cuando existen muchas posibilidades terapéuticas.
- b) En casos donde se desea evitar que animales enfermos escapen a la detección (ej.: enfermedades zoonóticas severas, enfermedades exóticas en un país o región).
- c) Cuando la probabilidad de hallar animales enfermos es baja.

Por su parte, las pruebas diagnósticas con alta especificidad son útiles para:

- a) Confirmar diagnósticos detectados mediante otras pruebas diagnósticas.
- b) Cuando los resultados falsos positivos tienen grandes efectos económicos y sociales.

Bibliografía

- Adler, H.E. and Wiggins, A.D. (1973).** Interpretation of serologic tests for *Mycoplasma gallisepticum*. *World. Poul. Sci.* **29**: 345–349.
- Galen, R.S. and Gambino, S.R. (1975).** Beyond Normality, the Predictive Value and Efficiency of Medical Diagnoses. John Wiley & Sons, New York, 237 pp.
- Miranda, A.O., Báez, E.N. y Tarabla H.D. (2002.)** Relación entre una modificación a la técnica del BPA y el valor predictivo positivo del diagnóstico de brucelosis bovina. *Vet. Arg.* **19**: 267–273.
- Houe, H., Kjaer Ersboll, A., Toft, N., Agger, J.F. (2003).** Veterinary epidemiology: from hypothesis to conclusion. Ed. The Royal Veterinary and Agricultural University, Frederiksberg, Dinamarca, 340 pp.
- Seiler, R.J. (1979).** The non-disease reactor: Considerations on the interpretation of screening test results. *Vet. Rec.* **105**: 226–228.
- Snedecor, G.W. and Cochran, W.G. (1980).** Statistical Methods. Iowa St. Univ. Press, 7th Ed., 507 pp.
- Yamamoto, R. (1983).** Factors affecting sensitivity and specificity of serologic tests: Some practical considerations and general thoughts. Univ. California Davis, EPM 216 Notes, 5 pp.
- Yerushalmy, J. (1947).** Statistical problems in assessing methods of medical diagnosis, with special reference to X-ray techniques. *Pub. Hlth. Rep.* **62**: 1432–1449.

Capítulo 6

Determinación de la sensibilidad y la especificidad de la prueba

1. Introducción

Hasta el momento hemos hablado de la sensibilidad y especificidad de las pruebas diagnósticas, su interpretación y los factores que las afectan. En este capítulo vamos a desarrollar la manera como se realizan los estudios tendientes a definir cuáles son la sensibilidad y especificidad de una prueba. Estos estudios suelen requerir diseños experimentales complejos y un análisis posterior de los datos que en algunas ocasiones excede el propósito de este libro. Por ello, desarrollaremos los aspectos básicos de los estudios para validar pruebas diagnósticas y los fundamentos estadísticos que soportan la determinación de la sensibilidad y especificidad de las pruebas.

Una prueba puede ser evaluada desde el punto de vista analítico midiendo el grado de acuerdo de la concentración de la sustancia de interés y la concentración verdadera. Una segunda opción es efectuar una evaluación diagnóstica, relacionando los resultados obtenidos con el verdadero estado de salud del individuo.

Por otra parte, el diseño de estudios para evaluar la exactitud de las pruebas diagnósticas puede ser considerado bajo dos puntos de vista: el primero considera aspectos de confiabilidad y el segundo de validez. Aunque ambos son elementos independientes de toda prueba, la confiabilidad es un importante prerrequisito para la validación de un proceso analítico (Duffy, 1994). La validación diagnóstica de una prueba considera todos los temas vinculados con la medición del error sistemático o la falta de él. Esto es, en definitiva, cuantificar la habilidad de la prueba en detectar individuos verdaderos positivos en la población enferma (sensibilidad) y verdaderos negativos en la población sana (especificidad) (Thorner y Remein, 1961).

Los estudios diseñados para estimar la sensibilidad y especificidad de las pruebas deben ser válidos tanto interna como externamente (Fleiss, 1981:321). En el primer caso, deben dar un estimador no sesgado del desempeño de la prueba bajo las condi-

ciones en las que se realizó el trabajo (protocolo, muestras de referencia y laboratorio). En el segundo, la muestra bajo estudio debe ser representativa de la población de referencia. Antes de diseñar un estudio para validar una prueba se debe tener muy en claro para qué va a ser utilizada la prueba (tamizado o confirmación, diagnóstico individual o grupal, etc.). El estudio puede ser experimental, observacional o mixto. Los ensayos experimentales son muy útiles para obtener resultados preliminares, pero generalmente tienden a sobrestimar el desempeño de la prueba. En todos los casos se deben seleccionar apropiadamente estrategias para reducir los sesgos (ej.: evaluaciones ciegas).

La determinación de la sensibilidad se efectúa en un grupo representativo de animales enfermos. En el caso de la especificidad, sin embargo, en la mayoría de los casos se necesita una mayor muestra representativa de los animales sanos. En la determinación de esta última no se deben utilizar aparentemente sanos provenientes de rodeos infectados, dado que ello llevará, generalmente, a subestimaciones de la especificidad de la prueba (Galen y Gambino, 1975:237).

Para determinar el tamaño de la muestra de animales a incluir en el estudio se usa la fórmula para estimar el intervalo de confianza de una proporción:

$$n = \frac{Z^2 \times P \times (1-P)}{e^2}$$

donde:

n= tamaño de la muestra para estimar la sensibilidad o especificidad de una prueba diagnóstica

Z= valor de la distribución Normal para un nivel de confianza especificado

P= sensibilidad o especificidad

e= error absoluto esperado (indica el grado de error que estamos dispuestos a aceptar en la estimación de la sensibilidad o la especificidad)

Supongamos que un estudio experimental preliminar había indicado que una determinada prueba de ELISA para la detección de una enfermedad tenía una sensibilidad y especificidad del 90%. Ahora se debe diseñar un estudio para estimar la sensibilidad y especificidad de la prueba de ELISA y tenemos que calcular el tamaño muestral con un margen de error del 5% y una confianza del 95%. ¿Cuál sería entonces el tamaño de la muestra que se requerirá?

$$n = \frac{(1,96)^2 \times 0,90 \times (1 - 0,90)}{(0,05)^2} = 138$$

De esta manera para nuestro estudio deberíamos tomar 138 individuos enfermos para la estimación de la sensibilidad y otros 138 individuos sanos para estimar la espe-

cificidad de la prueba de ELISA. Si los valores de sensibilidad y especificidad estimados previamente fuesen distintas (ej.: 90% de sensibilidad y 95% de especificidad), se debería considerar un número diferente de individuos sanos y enfermos para las respectivas estimaciones.

El error absoluto esperado (e) fue definido en el ejemplo anterior como del 5% (0,05), lo que indica que la estimación de la sensibilidad y especificidad puede variar entre 85% y 95% si el verdadero valor fuese del 90%. Si deseáramos reducir este error en la estimación y asumir no más del 2,5%, entonces deberíamos tomar 553 individuos tanto para la estimación de la sensibilidad como de la especificidad.

En términos generales el número de individuos a incluir en un diseño para la estimación de la sensibilidad y especificidad de una prueba diagnóstica se incrementa cuando: a) asumimos un menor error en la estimación (e), b) aumentamos la confianza en la estimación (Z) y c) el valor preliminar de sensibilidad o especificidad (P) se aleja del 100%.

Sin embargo, el cálculo del tamaño muestral tal como se lo definió anteriormente requiere de un conocimiento previo de los valores aproximados de sensibilidad y especificidad. En ausencia de estimaciones anteriores, en un primer ensayo se recomienda utilizar un mínimo de 100 individuos por grupo. Si la enfermedad es muy difundida o de gran importancia económica o impacto en salud pública, quizás un $n = 400$ por grupo sea un objetivo razonable (Gart y Buck, 1966).

El estado de salud de los animales a utilizar debe ser determinado con un alto grado de certeza por medio de pruebas biológicamente independientes de la que está siendo evaluada. Por otra parte, los animales sanos deben ser representativos de la población sana donde se aplicará la prueba (si se trata de una prueba indirecta para medir la cantidad de células somáticas en leche no se incluirán vacas recién paridas secretando calostro), mientras que en los animales enfermos debe estar presente todo el espectro clínico y patológico de la enfermedad en la misma proporción que en la población de referencia.

Las estimaciones de sensibilidad y especificidad se complican aún más en aquellas enfermedades multicausales que producen respuestas medibles de distinta magnitud de acuerdo con el patógeno actuante. Por ejemplo, en el caso de la mastitis bovina, la capacidad de reacción de la glándula mamaria para producir o alterar diversos elementos que pueden ser medidos por las pruebas diagnósticas y criterios de selección (Prueba de Mastitis de California o Antecedentes de Mastitis Clínicas) no es la misma ante cada especie bacteriana. Esto deriva en diferentes sensibilidades y especificidades para cada patógeno, que debe ser sumada a la variabilidad proveniente del animal (número y tiempo de lactancia), de la evolución de la enfermedad (si es una infección reciente o crónica), de la muestra de leche (ej.: primeros chorros antes del ordeño, durante el ordeño o leche residual) y de la prueba patrón (si se preincuban o no las muestras antes del cultivo bacteriano). Por otra parte, las estimaciones de

la especificidad de las pruebas diagnósticas para detectar infecciones intramamarias están influenciadas por el hecho de que los animales negativos provienen de rodeos que no son libres de infecciones intramamarias, por las variaciones fisiológicas en la composición de la leche relacionadas con el período de lactancia y por otros factores tales como el estrés ambiental (Gray y Martin, 1980).

A menudo, los investigadores están tentados de incluir solamente aquellos individuos que presentan pruebas claras de la enfermedad, tal como se miden con la prueba de referencia. A pesar de la certeza intelectual que parece proporcionar, esto puede redundar en una investigación que se limita a los individuos que tienen una enfermedad grave o en fase muy avanzada. Es importante preguntarse si con los enfermos estudiados se abarcó todo el espectro de la enfermedad (Riegelman y Hirsch, 1989:256). Un ejemplo de lo anterior fue resaltado por Vanasco et al. (2007) quienes desarrollaron una prueba tipo ELISA para el diagnóstico de leptospirosis en humanos y lo evaluaron en diferentes etapas de la enfermedad. La sensibilidad de esta prueba diagnóstica fue de 68,1 % cuando se aplicó a pacientes en la primer etapa de la enfermedad (menos de 10 días de evolución), aumentando a 93,2 % durante la segunda etapa (de 10 a 25 días de evolución) y volviendo a reducirse (78,8 %) en la tercera etapa (más de 25 días). Por otro lado, la especificidad de la prueba aumentó a medida que evolucionaba la enfermedad, logrando valores de 96,3, 99,3 y 100 % para la primera, segunda y tercera etapa, respectivamente.

De igual forma, dentro del grupo de individuos sanos se deberían incluir individuos que presenten enfermedades similares a aquella patología o condición que se desea diagnosticar. Con la mayor aceptación de la necesidad de validar las pruebas diagnósticas, aumenta paralelamente la importancia del diseño del estudio y la reducción del sesgo. Una evaluación sesgada produce típicamente una sobrestimación de la sensibilidad.

Si bien no existe una receta sobre cómo diseñar un estudio para calcular la sensibilidad y especificidad de una prueba diagnóstica, algunos de los pasos que deben seguirse son los siguientes:

1. Los investigadores seleccionan una prueba patrón que se usará para identificar individuos enfermos.
2. Seguidamente, escogen a un grupo de animales que según el criterio de referencia (prueba patrón) padecen la enfermedad y a otro grupo de individuos que según el mismo criterio están sanos. Al aplicar este criterio, los investigadores deben incluir a grupos representativos de individuos con y sin la enfermedad. En otras palabras, hay que asegurar que los individuos seleccionados representen el espectro completo de los que tienen la enfermedad y de los que no la tienen.
3. Los investigadores deben usar la prueba investigada para clasificar a todos los individuos como positivos o negativos.
4. Llegado a este punto, se ha podido clasificar a cada animal como sano o enfermo de acuerdo con la prueba patrón, y como positivo o negativo según el resultado de

la prueba a evaluar. Ahora ya pueden calcular el número de individuos en los que la prueba estudiada y la patrón concuerdan y en los que discrepan y presentarlos en una tabla de 2 x 2.

Es importante recordar que si bien la sensibilidad y la especificidad no están influidas directamente por la frecuencia relativa o prevalencia de la enfermedad, el número real de individuos que se identifican erróneamente como positivos o como negativos depende de la frecuencia relativa de la enfermedad.

2. Prueba patrón

Cuando se aplica cualquier prueba diagnóstica a animales que padecen una enfermedad o que no la padecen los resultados representan un recorrido de valores. En los enfermos, la variabilidad de los resultados puede reflejar diferencias en la gravedad de la enfermedad o una respuesta individual a la misma. A pesar de esta variabilidad, es esencial definir un grupo de animales que, sin lugar a dudas, padecen la enfermedad. La prueba o criterio utilizado para definir inequívocamente una enfermedad se conoce como prueba patrón o prueba de oro (*gold standard* por su término en inglés). El uso de un criterio de oro con el fin de identificar definitivamente a los que tienen la enfermedad es un requisito para examinar la utilidad diagnóstica de cualquier prueba nueva o no evaluada. En otras palabras, la utilidad de la nueva prueba se basa en su comparación con la patrón (Riegelman y Hirsch, 1989:256).

Puede parecer erróneo afirmar que la única forma de evaluar la capacidad diagnóstica de una prueba nueva es suponer que ya es posible realizar diagnósticos perfectos. Lamentablemente, esa es la posición en que nos encontramos al evaluar una prueba nueva. Sólo podemos preguntarnos si la prueba está a la altura de la mejor de las pruebas antiguas, esto es, la prueba patrón.

Por desgracia, incluso las mejores pruebas de referencia disponibles muchas veces no distinguen inequívocamente a los enfermos de los sanos. No existen muchas pruebas patrón, en algunos casos puede llegar a ser una de «bronce» o en el peor de los casos hasta una de «acero» (Dohoo et al., 2003:706).

La prueba patrón suele ser la más riesgosa, la más difícil técnicamente, la más cara o la opción diagnóstica más impráctica dentro de las disponibles, incluso, para muchos padecimientos no existe prueba patrón disponible (Knapp y Clinton Miller, 1992:435). En la mayoría de las enfermedades infecciosas la prueba patrón es la identificación de los patógenos mediante su cultivo y aislamiento (Nielsen et al., 1992:203). Sin embargo, en muchas ocasiones este es un proceso muy lento y costoso, mientras que en otras se agrega el inconveniente que no es posible aislar fácilmente al organismo responsable de la enfermedad (ej.: brucelosis, enfermedades virales). En otros casos, la prueba patrón puede ser una biopsia, una necropsia o cualquier criterio que pueda diagnosticar inequívocamente la enfermedad.

En muchas ocasiones, la prueba patrón puede ser en realidad una combinación de pruebas diagnósticas. Por ejemplo, la definición de caso de leptospirosis humana puede no agotarse con la sola positividad a la prueba de microaglutinación en tubo (MAT), sino que puede combinar a lo anterior (seroconversión en la MAT en dos o más muestras con al menos una semana de intervalo), la presencia de neutrofilia mayor de 70 %, un número de leucocitos mayor de 8000/mm³ y que los pacientes presenten antecedentes clínicos y epidemiológicos indicativos de leptospirosis (Vanasco et al., 2007).

3. Relación entre sensibilidad y especificidad de la prueba y la prevalencia de la enfermedad

En general, la sensibilidad y especificidad no varían al cambiar la prevalencia real de la enfermedad, siempre que dicho cambio no altere la distribución de los atributos de la enfermedad en la población. Cualquier cambio que afecte la forma de las curvas de individuos enfermos y/o sanos o el rango de superposición entre ambas cambiará los patrones de sensibilidad y especificidad de la prueba. Si la relación con la prueba diagnóstica es importante, estos cambios incluirán la distribución de razas, edad, sexo, exposición a afecciones antigénicamente relacionadas (Vanasco et al., 2007). Estas variaciones en la población deben ser tenidas en cuenta pero no pueden ser controladas (Yamamoto, 1983:5). Por razones prácticas se asume que la sensibilidad y la especificidad de la prueba son independientes de la prevalencia verdadera y de las variaciones existentes entre rodeos, áreas geográficas y otras características demográficas. Sin embargo, no se debe olvidar que en un programa de erradicación de enfermedades que incluya eliminación de los animales reaccionantes positivos se eliminan con ellos los positivos verdaderos (enfermos) pero también los falsos positivos. De esta manera, la prevalencia de reaccionantes cruzados en la población sana puede disminuir y ser menor a la utilizada cuando se calculó la especificidad de la prueba. En la práctica, esto conlleva a un aumento en la especificidad de la prueba a medida que el programa progresa. Por ello, se debe recolectar información sobre las características demográficas de los animales que están sujetos a evaluación para revisar estos supuestos cada vez que sea posible (Martin, 1984). Por último, se debe tener en cuenta que algunos aspectos de la determinación de la sensibilidad y especificidad de una prueba (ej.: si la prueba se usa repetidamente en los mismos animales) se relacionan no sólo con el poder discriminatorio de la prueba diagnóstica sino también con el poder discriminatorio del programa de vigilancia epidemiológica en su conjunto (Gray y Martin, 1980).

4. Determinación de la concordancia entre pruebas diagnósticas

En la práctica es muy difícil cumplir con todas las premisas para determinar sensibilidad y especificidad, por lo que se acostumbra a medir los resultados de la nueva prueba comparándolos con los de otra(s) prueba(s) que no es(son) biológicamente independiente(s) de la que está siendo evaluada. Por ejemplo, si se tienen dos pruebas disponibles, se tomarán como enfermos los positivos a las dos pruebas y como sanos los negativos a ambas. Sin embargo, se debe recordar que estas estimaciones, aunque útiles, pueden estar sesgadas dado que es improbable que nuestras muestras de población sana y enferma incluyan el espectro completo de las poblaciones de animales sanos y enfermos. Además, si las pruebas que se quieren comparar están biológicamente relacionadas, las mismas tenderán a dar los mismos resultados al ser aplicadas en un mismo individuo. Por todo ello, en estas circunstancias es conveniente llamar a nuestras estimaciones como sensibilidad y especificidad relativas (Hanley y McNeil, 1982) y a la prueba en la cual se basó el diagnóstico como prueba de referencia (Briesofsky y Greiner, 1998). Estas comparaciones sólo serán valiosas si la sensibilidad y especificidad de esta prueba de referencia se aproximan al 100 % (Cannon y Roe, 1982:35). Cuando no se cuentan con pruebas definitivas para determinar el estado de salud verdadero de los animales, la sensibilidad y la especificidad pueden ser estimadas por métodos estadísticos especiales (Martin et al., 1992). Estas estimaciones, sin embargo, pueden ser sesgadas en el caso de pruebas biológicamente asociadas (Cohen, 1960).

Se debe tener en cuenta que la comparación de dos o más pruebas sólo indica el grado de concordancia entre los resultados de las pruebas, pero de ninguna manera si una prueba es más sensible o específica que otra (Greiner y Gardner, 2000). Además, estas comparaciones deben hacerse con un diseño ciego, de manera que la persona que efectúa una prueba desconozca el resultado de la otra prueba para evitar serios sesgos. Cuando se comparan dos pruebas diagnósticas aplicadas en poblaciones diferentes o en diferentes subgrupos de la misma población, los índices de concordancia y la probabilidad de error pueden variar de acuerdo con la prevalencia real de la enfermedad en las poblaciones o subpoblaciones sujetas a comparación (Elizalde et al., 2009; Gardner y Greiner, 1999:74).

La manera más simple y frecuente para medir el grado de acuerdo entre los resultados de dos pruebas (A y B) es la proporción de concordancia general o Concordancia Observada (CO). Ésta es igual a la suma de la probabilidad de los resultados positivos y negativos a ambas pruebas (Donald et al., 1994).

$$\text{Concordancia Observada} = [\text{Pr A (+) y B (+)}] + [\text{Pr A (-) y B (-)}]$$

Sin embargo, se necesita más que una medida del grado de acuerdo entre dos pruebas (Concordancia Observada), dado que, salvo en circunstancias extremas, es dable esperar algún grado de concordancia debido solamente a la intervención del

azar. Una medida que corrige esto es la razón denominada Kappa (propuesto por Cohen en 1960), que incluye el cálculo de la Concordancia Esperada (CE) (Buck y Gart, 1966). La CE incorpora los valores marginales de cada prueba y es igual a:

$$\text{Concordancia Esperada} = \frac{\text{Pr A (+)} \times \text{B (+)} + \text{Pr A (-)} \times \text{B (-)}}{n^2}$$

Por su parte, Kappa es igual a:

$$K = \frac{CO - CE}{1 - CE}$$

Para entender el índice Kappa nos haremos dos preguntas. Primero, ¿cuánto mejor es la coincidencia entre observadores con respecto a aquella que podríamos esperar sólo por azar? Esto se calcula restando el porcentaje de coincidencia observado entre los observadores menos el porcentaje de coincidencia esperado solo por azar. Éste es el numerador del índice Kappa.

La segunda pregunta es ¿cuál es la máxima coincidencia entre dos observadores que podríamos tener sobre la coincidencia sólo por azar? Claramente la máxima coincidencia sería del 100 %. Por lo tanto, la máxima coincidencia esperada sólo por azar sería de 100 %, siendo éste el denominador del índice Kappa.

Entonces, este índice expresa el grado de coincidencia que excede a la que se espera por azar (numerador) con relación a la máxima coincidencia esperada más allá del azar (denominador).

En el ejemplo del Cuadro 1 calculemos Kappa.

Cuadro 1. Cuadro de cálculo de la concordancia entre dos pruebas diagnósticas en una población de 1000 animales

Condición	Test B positivos	Test B negativos	Total
Test A positivos	90	100	190
Test A negativos	10	800	810
Total	100	900	1000

La Concordancia Observada es equivalente a:

$$CO = \frac{90}{1000} + \frac{800}{1000} = \frac{90 + 800}{1000} = 0,89$$

y la Concordancia Esperada:

$$CE = \frac{190 \times 100 + 810 \times 900}{1000^2} = 0,75$$

Kappa es entonces:

$$K = \frac{0,89 - 0,75}{1 - 0,75} = 0,56$$

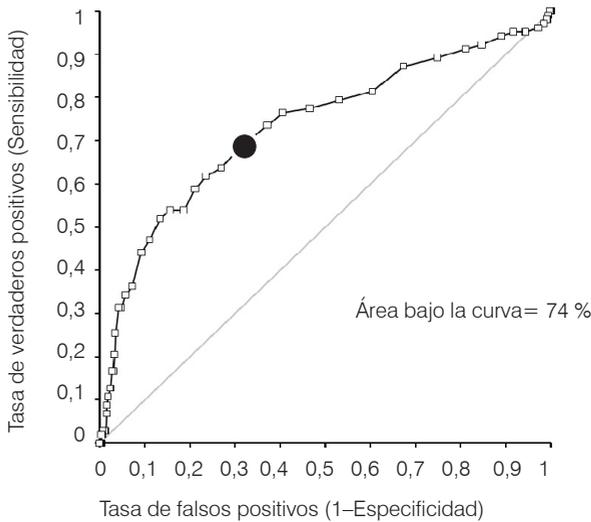
Landis y Koch (1977) sugirieron que un índice Kappa= 1 representa una concordancia perfecta, más allá de la intervención de azar, K = 0 no hay concordancia. Para la mayoría de los casos, valores mayores a 0,75 se interpretan como una concordancia excelente, entre 0,75 y 0,40 como buena a regular y menores a 0,40 como pobre (Gray y Martin, 1980). En el caso de la comparación entre pruebas diagnósticas, se necesita un Kappa de al menos 0,40 a 0,50 para indicar un nivel de acuerdo moderado (Hui y Walter, 1980).

5. Curva de la característica operativa del operador

Para el cálculo de las áreas de superposición en las pruebas que expresan sus resultados bajo una escala continua puede utilizarse una técnica de análisis de una curva conocida generalmente por su sigla en inglés ROC (de *Receiver Operating Characteristic*) (Galen y Gambino, 1975:237). Esta curva es una representación gráfica de la relación entre la sensibilidad y la especificidad de una prueba diagnóstica. Se construye graficando la sensibilidad contra (1 – especificidad) con varias posibilidades de criterios de positividad y describe la habilidad de una prueba para discriminar entre animales enfermos y sanos (Buck y Gart, 1966; Gart y Buck, 1966).

El área bajo la curva es el parámetro estadístico más importante que genera la curva ROC y es igual a la probabilidad de que un individuo con la enfermedad tomado al azar tenga un valor mayor en la medición que estamos efectuando que un individuo al azar sin la enfermedad. Una prueba perfecta tiene un área del 100 %, mientras que una prueba que no mejora lo que obtendríamos por azar tiene una área del 50 % (Briesofsky y Greiner, 1998) limitada por una línea recta diagonal desde el ángulo inferior izquierdo al superior derecho (Figura 1.). Si bien no existe una guía definida, se puede distinguir entre pruebas con nula capacidad discriminatoria cuando su área bajo la curva es de 0,5, y pruebas con baja (áreas entre 0,5 y 0,7), moderada (áreas entre 0,7 y 0,9) y alta capacidad discriminatoria (áreas entre 0,9 y 1).

Figura 1. Datos de una curva ROC de una prueba de conductividad eléctrica para determinar la presencia de infecciones intramamarias en vacas lecheras (Elizalde et al., 2009)



Referencias: ● Punto de corte óptimo (independientemente del costo de los FN y FP)

La curva ROC es una herramienta simple para la aplicación del método diagnóstico o del valor predictivo mencionado en el punto 1.2.2. del capítulo 3 en la selección del criterio de «anormalidad» (Gart y Buck, 1966) y puede ser calculada por medio de programas de computación como el ROC Analysis (Metz, 1998) y el MedCalc (Schoonjas et al., 1995). El procedimiento TG-ROC (del inglés *Two-graph ROC*) permite, además, leer las mediciones de exactitud del gráfico directamente para un punto de corte seleccionado y definir resultados intermedios de la prueba (Branscum et al., 2005).

En el gráfico anterior, si la importancia de los diagnósticos falsos positivos es igual a la de los falsos negativos, el punto de corte óptimo está dado por el valor que se señala con el círculo negro, ya que es el punto que más se acerca al ángulo superior izquierdo y que se corresponde con una sensibilidad del 70 % y una especificidad del 68 %. Sin embargo, en los casos en que el costo de efectuar un diagnóstico falso positivo es diferente del de uno falso negativo, este punto de corte puede variar aumentando la especificidad o la sensibilidad respectivamente (Knapp y Clinton Miller, 1992:435). Esta curva también puede ser utilizada para comparar dos pruebas diagnósticas (Vanasco et al., 2011). La mejor prueba será aquella cuya área bajo la curva sea mayor (Gart y Buck, 1966).

6. Valores de puntos de corte

Como ya se ha visto en capítulos anteriores, existe un solapamiento entre los resultados de las pruebas diagnósticas aplicadas a individuos sanos y enfermos, lo que torna imposible seleccionar un punto de corte que permita discriminar perfectamente entre individuos sanos y enfermos (Ver Figuras 3 y 4 del capítulo anterior). Por ello, el punto de corte sólo discrimina entre individuos positivos o negativos a la prueba, englobando en esas categorías individuos verdaderamente sanos o enfermos y falsamente clasificados como portadores o no de la enfermedad.

La determinación del valor de punto de corte es aplicable para pruebas cuyos resultados se expresan en escalas ordinales, conteos o continuas. Lamentablemente no existe una única manera de seleccionar el punto de corte ideal para cada situación en particular.

6.1. Selección del punto de corte para maximizar la sensibilidad de la prueba

En algunas ocasiones es deseable reducir al mínimo la probabilidad de resultados falsos negativos, por lo que se tiende a seleccionar bajos puntos de corte, lo cual torna a la prueba altamente sensible. Las situaciones más apropiadas para seleccionar esta alternativa son:

- a)** En aquellos casos donde la presencia de individuos portadores de la enfermedad se considera extremadamente peligrosa para la sanidad animal o la salud pública (ej.: animales con encefalopatía espongiforme bovina o aftosa).
- b)** Cuando obtener resultados falsos positivos no es de gravedad ya que los mismos pueden ser reanalizados con otra prueba complementaria de mayor especificidad.
- c)** Esta situación se da cuando existe una secuencia de análisis, evaluando inicialmente a los animales mediante una prueba tamiz (ej.: BPA para brucelosis) y a los positivos se los analiza mediante una prueba complementaria (ej.: fijación de complemento).
- d)** Cuando los individuos enfermos pueden ser tratados con buenos resultados, mientras que los que no son tratados tienen un desenlace fatal.

6.2. Selección del punto de corte para maximizar la especificidad de la prueba

En el otro extremo, puede estar justificado seleccionar el punto de corte de forma tal de reducir la proporción de diagnósticos falsos positivos a expensas de diagnósticos falsos negativos, incrementando el valor del punto de corte. Es factible seleccionar un punto de corte que maximice la especificidad de la prueba si en paralelo se realiza otra prueba con mejor sensibilidad o en aquellos casos donde diagnosticar falsos positivos tenga un costo demasiado alto.

Como se observa, utilizando este criterio se incrementará la proporción de resultados falsos negativos, por lo que su uso está restringido sólo a aquellos casos en los que las consecuencias asociadas a este tipo de resultados no sean muy severas. No obstante, diagnosticar como negativo a un animal que en realidad está enfermo puede ser desventajoso en términos de costos y de bienestar animal.

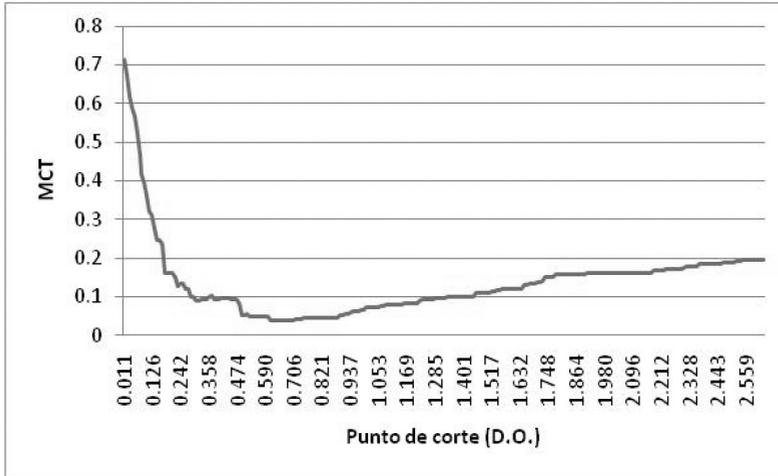
6.3. Selección del punto de corte en función de los costos

En algunas ocasiones es posible contar con una aproximación de la prevalencia de la enfermedad en la población y los costos asociados con una clasificación errónea tanto de animales falsos negativos como falsos positivos. En estos casos, se puede emplear el concepto introducido por Berkson en 1947 sobre la selección de puntos de corte en términos de costos por clasificación errónea de los animales (MCT, por «*Misclassification Costs Term*»). Bajo esta óptica se deben considerar entonces los costos monetarios o médicos, sociales u otras desventajas asociadas con los diagnósticos falsos (Vanasco et al., 2007). El MCT se calcula de la siguiente forma:

$$\text{MCT} = (1 - \text{Prevalencia}) \times (1 - \text{Especificidad}) + r \times \text{Prevalencia} \times (1 - \text{Sensibilidad})$$

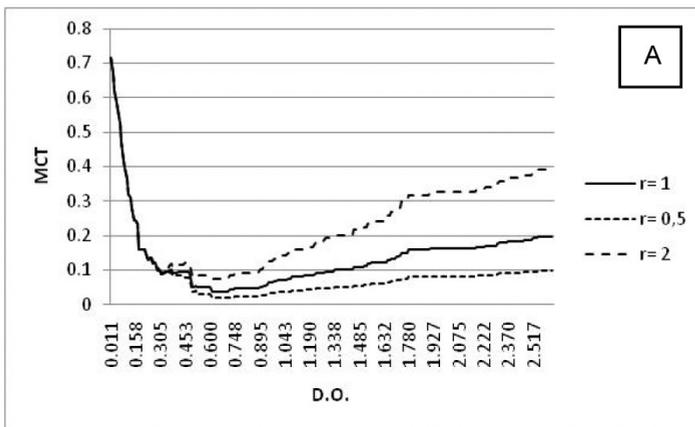
donde r es la razón de los costos atribuibles a los diagnósticos falsos negativos y falsos positivos, respectivamente ($r = \text{Costos de los Falsos Negativos} / \text{Costos Falsos Positivos}$). El punto de corte seleccionado es aquel que genera el menor valor de MCT. Existe un paquete informático (TG-ROC) que genera un gráfico donde se visualizan los valores de MCT en función de los diferentes puntos de corte. En la Figura 2 se presenta un caso hipotético de una prueba de ELISA (unidad de medida en densidad óptica, DO) para una enfermedad de ovinos, la cual tiene una prevalencia del 20 % y se asume que los costos de los resultados falsos negativos son iguales a los de los diagnósticos falsos positivos ($r = 1$).

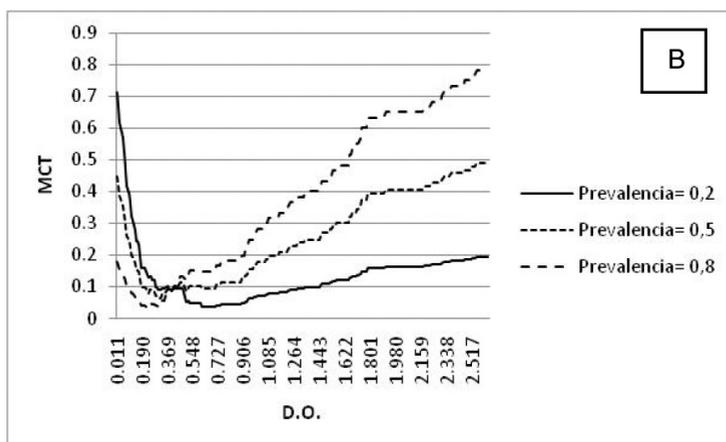
Figura 2. Selección del punto de corte en función del costo de los diagnósticos falsos negativos y falsos positivos (prevalencia= 0,2; $r= 1$)



Como se ve en la figura anterior, el punto de corte óptimo en este caso se encuentra entre valores de densidad óptica de 0,611 y 0,706. No obstante, si modificamos el valor de r para priorizar los resultados falsos negativos ($r= 2$) o los falsos positivos ($r= 0,5$), podemos observar (Figura 3, A) que el punto de corte, si bien no varía significativamente del calculado para el caso en el cual ambos resultados falsos tenían la misma preponderancia ($r= 1$), tiende a incrementarse a medida que el costo de los falsos positivos se torna más influyente. Por otro lado, cuando modificamos el dato de la prevalencia de la enfermedad en la población se advierte (Figura 3, B) que el punto de corte disminuye a medida que la prevalencia aumenta.

Figura 3. Comparación de la selección del punto de corte en función de la razón de costos entre falsos negativos y falsos positivos (A) y la prevalencia de la enfermedad en la población (B)





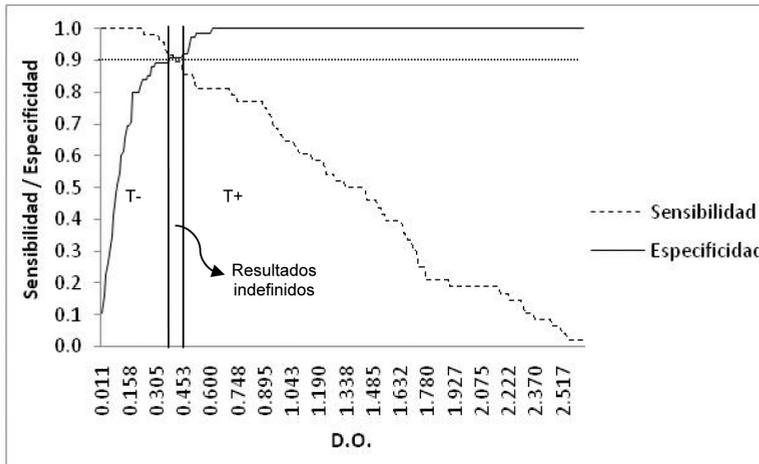
Aunque esta metodología resulta intuitivamente apropiada, no siempre es factible contar con la información necesaria para considerar los costos de diagnosticar incorrectamente animales negativos o positivos.

6.4. Selección de puntos de corte intermedios

Cuando el resultado de una prueba diagnóstica resulta alejado del punto de corte, independientemente de si el valor es menor o mayor al mismo, el veterinario puede estar razonablemente seguro de la condición del animal. Pero cuando el resultado da un valor cercano al punto de corte, es lógico dudar sobre la clasificación del estatus sanitario de ese animal. Cuando sucede esto último se dice que el resultado es «indefinido» y el individuo es catalogado como «sospechoso» dado que cae en una «zona borrosa o gris» que dificulta su interpretación.

En este tipo de escenarios es ventajoso establecer tres tipos de resultados: positivos, negativos e intermedios. Por ende, ya no tenemos que definir un punto de corte sino dos. Estos puntos de corte se definen usualmente con base en un valor de confianza para la sensibilidad y especificidad que se establezca previamente (normalmente 90 %). La principal desventaja de esta metodología es que, en determinadas situaciones y según el tipo de prueba, esta zona gris puede ocupar una gran parte del espectro de resultados y queda así una estrecha zona remanente en la cual los resultados son claramente negativos o positivos (Figura 4).

Figura 4. Selección de dos puntos de corte en función de la confianza (90 %) de la sensibilidad y especificidad



Aquellos animales cuyos resultados caigan en la zona gris, podrán ser posteriormente evaluados con la misma u otra técnica para definir su estatus sanitario. Un ejemplo típico de estas interpretaciones es el que se da en los antibiogramas. En muchas ocasiones existen tres resultados posibles: resistente al antimicrobiano, sensible al antimicrobiano o intermedio.

6.5. Selección del punto de corte en función de los factores de riesgo

Como ya lo hemos expresado en varias oportunidades, los animales responden a una misma enfermedad generando diferentes respuestas derivadas de la propia variación biológica. Este abanico de respuestas explica por qué no hay pruebas que sean 100 % sensibles y específicas al mismo tiempo. De tal forma se pueden identificar algunos factores que afectan la respuesta del animal ante una enfermedad, tal es como el sexo, la edad, la raza y otros factores de tipo ambiental (ej.: condiciones de manejo, clima). En aquellas ocasiones donde el investigador haya desarrollado un estudio tendiente a validar una técnica diagnóstica y haya consignado al mismo tiempo la exposición a determinados factores asociados con la respuesta animal, se puede desarrollar una regresión logística. Esta técnica estadística permite medir el impacto que cada factor tiene sobre la respuesta medible del animal.

Supongamos que estamos tratando de validar una técnica de ELISA para el diagnóstico de leptospirosis en bovinos. Tomamos una población de individuos sanos y otra de enfermos (empleando la técnica de Microaglutinación en Tubo —MAT— como prueba de referencia) y paralelamente tomamos datos sobre la edad de los animales, el sexo, la raza, las condiciones de manejo, algunos parámetros ambientales (humedad,

presencia de roedores, etc.). Posteriormente realizamos una regresión logística, la cual nos genera el siguiente modelo:

$$\ln(p/1-p) = \beta_0 + \beta_e X_e + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

donde $p = P(E+)$

β_0 = intercepción del modelo (similar a la ordenada al origen en regresiones lineales)

β_e = coeficiente para el valor de ELISA

X_e = valor de ELISA observado

β_i = coeficiente para cada uno de los factores biológicos estudiados ($i = 1, 2, \dots, k$)

X_i = valor observado para el i -ésimo factor biológico estudiado.

De este modo, se combina el resultado de la prueba de ELISA con las características biológicas del animal para dar una estimación de la probabilidad de que ese animal esté verdaderamente enfermo.

6.6. Selección del punto de corte de acuerdo con el Índice de Youden (J)

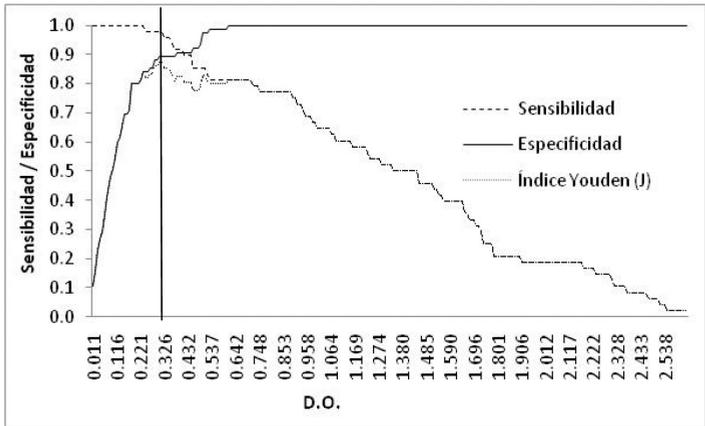
Este índice es otra medida independiente de la prevalencia de la enfermedad que utiliza la sensibilidad y la especificidad de la prueba para cada punto de corte posible. Se calcula de la siguiente forma:

$$\text{Índice de Youden (J)} = \text{Sensibilidad} + \text{Especificidad} - 1$$

Básicamente este índice refleja la diferencia entre la tasa de verdaderos positivos y la de falsos positivos. Un buen punto de corte para la prueba diagnóstica es aquel que resulte tener alta esta diferencia. En otras palabras, el mejor punto de corte estimado mediante J será aquel que más se aproxime a 1 (en teoría, la prueba perfecta es aquella que ofrezca un valor de $J=1$, dado que así la sensibilidad y la especificidad de la prueba serían perfectas). El índice de Youden tiene la ventaja de no estar afectado por la prevalencia de la enfermedad y es uno de los parámetros preferidos puesto que utiliza una combinación simple de los valores de sensibilidad y especificidad. Sin embargo, tiene la desventaja de que, al resultar de la combinación de los valores de sensibilidad y especificidad, no es posible determinar si la prueba diagnóstica es buena en sensibilidad, especificidad o en ambas.

Siguiendo con los ejemplos anteriores, en la Figura 5 se muestran los mismos datos empleados para validar la técnica de ELISA y el Índice de Youden para cada valor de corte posible. El valor de DO que maximiza el índice es el propuesto como punto de corte, y en nuestro ejemplo es el valor correspondiente a una $DO = 0,316$, y de esta manera la prueba diagnóstica tendría una sensibilidad = 0,978 y una especificidad = 0,896.

Figura 5. Selección del punto de corte con base en el Índice de Youden



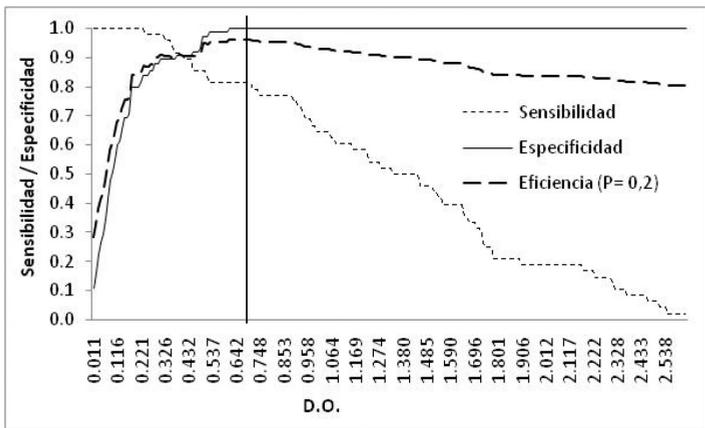
6.7. Selección del punto de corte de acuerdo con la eficiencia

Otro parámetro que puede ser de utilidad para definir el punto de corte es la eficiencia, la cual considera la prevalencia de la enfermedad en el cálculo:

$$\text{Eficiencia} = \text{Prevalencia} \times \text{Sensibilidad} + (1 - \text{Prevalencia}) \times \text{Especificidad}$$

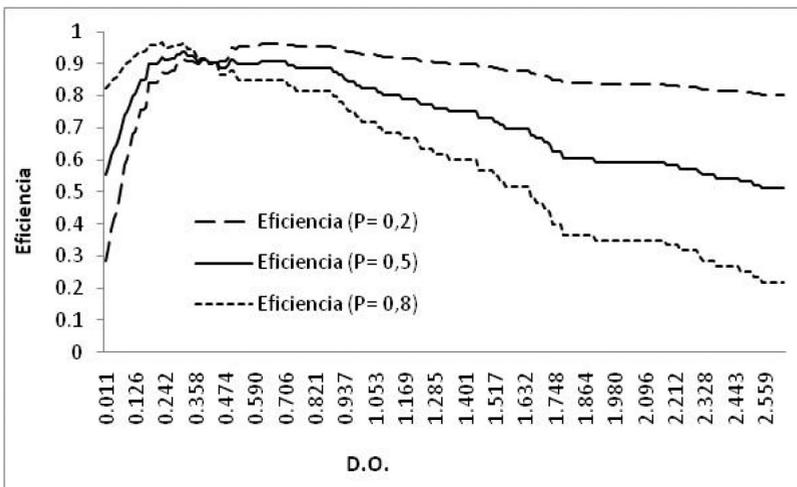
En la Figura 6. se observan, para la validación de la técnica de ELISA, la sensibilidad, la especificidad y la eficiencia de acuerdo con el punto de corte seleccionado. El valor de ELISA que maximice la eficiencia será el seleccionado. En el ejemplo expuesto, el punto de corte sería el valor de 0,727 (considerando una prevalencia del 20 %), el cual genera que la prueba diagnóstica tenga una sensibilidad de 0,815 y una especificidad de 1.

Figura 6. Selección del punto de corte con base en la eficiencia



La principal desventaja de esta metodología radica en que, dado que la eficacia es un parámetro dependiente de la prevalencia, esto obliga a modificar el punto de corte a medida que varía la prevalencia de la enfermedad en la población. En la Figura 7 se muestran las diferentes curvas de eficiencia según la prevalencia sea del 20 %, 50 % u 80 %. De igual manera, los puntos de corte que maximizan la eficiencia también son diferentes. Para una prevalencia del 20 %, el punto de corte era un valor de $DO = 0,727$ con una sensibilidad = 0,815 y una especificidad = 1. Cuando se simuló una prevalencia del 50 %, el punto de corte óptimo correspondió al valor de $DO = 0,31$, con sensibilidad y especificidad de 0,98 y 0,89, respectivamente. Cuando se consideró una prevalencia del 80 % el punto de corte óptimo fue un valor de $DO = 0,232$, y generó una sensibilidad de 0,995 y una especificidad de 0,89. Es decir, a medida que la prevalencia de la enfermedad aumenta el punto de corte es menor, lo que hace que la prueba gane en sensibilidad y pierda especificidad.

Figura 7. Selección del punto de corte con base en la eficiencia según la prevalencia de la enfermedad



7. Estimación de la sensibilidad y especificidad en ausencia de prueba patrón

En la mayoría de los casos, las pruebas existentes no tienen las características suficientes para ser calificadas como pruebas patrón y por lo tanto queda en duda su utilidad para servir como base en el cálculo de la sensibilidad y especificidad de una nueva prueba diagnóstica que desee ser evaluada. Si no existe una prueba patrón adecuada, una de las posibilidades es medir el grado de concordancia entre la nueva prueba y la existente hasta el momento aplicando el test Kappa que ya fue tratado anteriormente en este capítulo.

Una alternativa a este test Kappa, especialmente para pruebas con resultados de tipo cuantitativos, es estimar la sensibilidad y especificidad de la nueva prueba a través de métodos bayesianos. El tratamiento profundo de estas herramientas escapa a los objetivos de este libro, por lo que aquí trataremos los conceptos generales. Para aquellos lectores que posean un interés particular en el desarrollo de estas metodologías, recomendamos la lectura de los siguientes artículos: Pouillot y col. (2002), Branscum y col. (2005) y Greiner y Gardner (2000).

7.1. Disponibilidad de la sensibilidad y especificidad de la prueba patrón

En algunas ocasiones podemos contar con información sobre la sensibilidad y especificidad de una prueba que, sin ser considerada una prueba patrón, es la que se emplea como referencia ($Sens_{ref}$ y Esp_{ref} , respectivamente) para el diagnóstico de una enfermedad. En esta circunstancia, es posible construir una tabla de 2 x 2 basada en los resultados de la nueva prueba (pero con el estatus de enfermedad determinado a partir de los resultados de la prueba de referencia) y estimar la $Sens_{nueva}$ y la Esp_{nueva} de la nueva prueba usando la sintaxis de la siguiente manera:

$$Sens_{nueva} = \frac{T_{nueva}^+ \times Esp_{ref} - (T_{nueva}^+ \text{ y } T_{ref}^-)}{n \times Esp_{ref} - T_{ref}^-}$$

$$Esp_{nueva} = \frac{T_{nueva}^- \times Sens_{ref} - (T_{ref}^+ \text{ y } T_{nueva}^-)}{n \times Sens_{ref} - T_{ref}^+}$$

Supongamos que estamos tratando de validar una prueba de ELISA para el diagnóstico de leptospirosis en bovinos. Como técnica de referencia empleamos Microaglutinación en Tubo (MAT) ya que puede considerarse que la misma no reúne todas las características que debería poseer una prueba patrón. Se analizan 535 muestras y los resultados de ambas pruebas se presentan en el Cuadro 2.

Cuadro 2. Evaluación de la técnica de ELISA para el diagnóstico de leptospirosis en bovinos empleando la técnica MAT como prueba de referencia ($Sens. = 95 \%$, $Esp. = 99 \%$)

Condición	ELISA positivos	ELISA negativos	Total
MAT positivos	354	2	356
MAT negativos	7	172	179
Total	361	174	535

Por lo tanto, tenemos que:

$$Sens_{ELISA} = \frac{T_{ELISA}^+ \times Esp_{MAT} - (T_{ELISA}^+ \text{ y } T_{MAT}^-)}{n \times Esp_{MAT} - T_{MAT}^-}$$

$$Sens_{ELISA} = \frac{361 \times 0,99 - 7}{535 \times 0,99 - 179}$$

$$\text{Sens}_{\text{ELISA}} = 0,999 \text{ o } 99,9 \%$$

$$\text{Esp}_{\text{ELISA}} = \frac{T_{-\text{ELISA}} \times \text{Sens}_{\text{MAT}} - (T_{+\text{MAT}} \text{ y } T_{-\text{ELISA}})}{n \times \text{Sens}_{\text{MAT}} - T_{+\text{MAT}}}$$

$$\text{Esp}_{\text{ELISA}} = 174 \times 0,95 - 2 / 535 \times 0,95 - 356$$

$$\text{Esp}_{\text{ELISA}} = 1,00 \text{ o } 100 \%$$

También se puede estimar la prevalencia

$$P = n \times (\text{Esp}_{\text{ref}} - 1) + T_{+\text{ref}} / n \times (\text{Sens}_{\text{ref}} + \text{Esp}_{\text{ref}} - 1)$$

$$P = 535 \times (0,99 - 1) + 356 / 535 \times (0,95 + 0,99 - 1)$$

$$P = 0,69 \text{ o } 69 \%$$

7.2. Estimación mediante los Coeficientes de Máxima Verosimilitud (*Maximum likelihood*)

Se puede estimar la sensibilidad y especificidad de dos o más pruebas diagnósticas a partir de datos obtenidos de dos poblaciones animales con diferentes prevalencias de la enfermedad en las cuales se hayan usado dichas pruebas diagnósticas.

A partir de estos datos se construyen dos tablas 2 x 2, lo que representa 6 grados de libertad. Son 6 los parámetros a calcular, la sensibilidad y especificidad y la prevalencia para cada una de las 2 poblaciones analizadas.

Este procedimiento está basado en tres supuestos críticos:

- a) Las pruebas diagnósticas tienen que ser independientes (no debe existir dependencia condicional entre los resultados de ambas pruebas).
- b) La sensibilidad y especificidad deben ser constantes a lo largo de todas las poblaciones evaluadas.
- c) La prevalencia de la enfermedad en las dos poblaciones debe ser diferente.

Cualquier violación a estos supuestos invalida la técnica.

La metodología se basa en el uso de los algoritmos de Newton–Raphson o un algoritmo EM. Se ha desarrollado un software para la estimación de los parámetros (TAGS), por lo que para ampliar este punto se recomienda la lectura del artículo de Pouillot et al. (2002).

7.3. Método bayesiano

El método bayesiano ofrece una aproximación más flexible para la estimación de la sensibilidad y especificidad en ausencia de una prueba patrón. Las estimaciones bayesianas pueden incorporar estimados *a priori* de la sensibilidad y especificidad de una prueba dentro del proceso. Ellos pueden también ser usados para relajar los requerimientos de datos de múltiples poblaciones o para construir factores que participan de dependencias condicionales a lo largo de los resultados de las pruebas.

No obstante, la discusión sobre estos procedimientos va más allá de los objetivos perseguidos por este libro, recomendándose la lectura de Branscum y col. (2005) y Greiner y Gardner (1999).

7.4. Meta-análisis

Para algunas pruebas diagnósticas es posible que se hayan realizado diferentes estudios de validación aplicados en diversas situaciones y en una gran variedad de individuos. En tanto, los resultados suelen ser muy variables entre los distintos estudios, situación que dificulta la obtención de conclusiones acerca de la eficacia global de esta prueba. Con el objetivo de obtener estimados globales de sensibilidad y especificidad, es posible realizar un meta-análisis que ofrezca un valor global para cada parámetro. La gran fortaleza de esta metodología estadística radica en que incorpora diferentes estudios de validación, cada uno conducido bajo diferentes condiciones, tipos de animales, razas, sexo, edades y condiciones ambientales y de manejo. Además, si se cuenta con un número importante de estudios, el meta-análisis puede detectar fuentes de variabilidad que se deban, entre otras cosas, a:

- a)** Particularidades de los hospedadores (ej.: raza, sexo, edad, fase de la enfermedad).
- b)** Detalles técnicos o adaptaciones de las pruebas de acuerdo con las condiciones de los diferentes laboratorios.
- c)** Procedimiento de toma de muestras.
- d)** Diseño de los estudios (ej.: ensayos ciegos, estratificados).

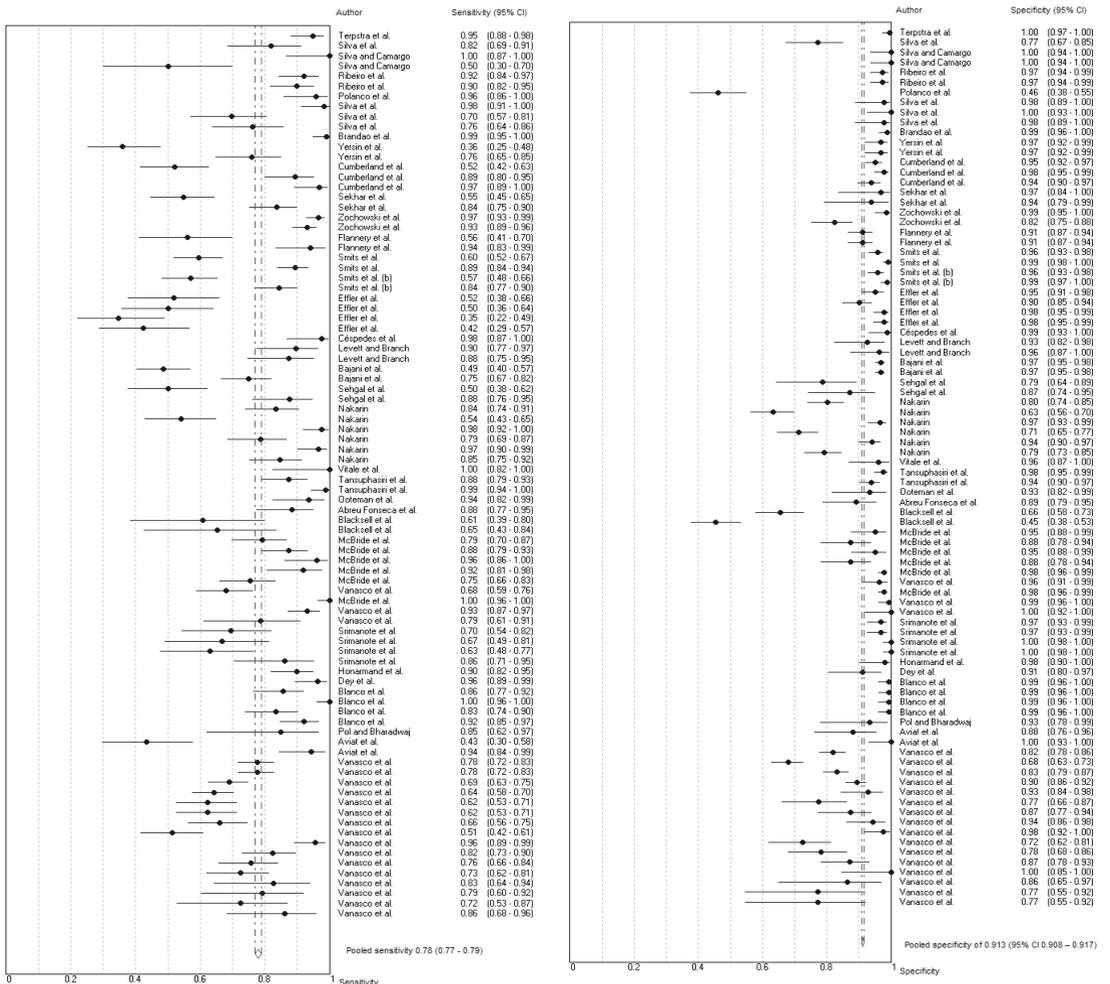
Para explicar esta metodología, tomaremos como ejemplo un meta-análisis conducido con el objetivo de evaluar la eficacia de las pruebas de ELISA para la detección de anticuerpos específicos contra *Leptospira* en humanos (Signorini et al., 2013). No se pretende aquí profundizar en los cálculos matemáticos que se realizan en un meta-análisis sino explicar brevemente los pasos básicos de esta metodología.

El meta-análisis es un proceso en dos etapas. En un primer paso se calculan los estadísticos de resumen que describen la validez de la prueba diagnóstica (sensibilidad y especificidad). En un segundo paso se calculan los índices globales como promedios ponderados de los índices individuales. El meta-análisis debería realizarse sólo si los estudios han sido conducidos con pacientes clínicamente similares, han evaluado pruebas comparables y han usado pruebas de referencia también comparables.

Inicialmente se seleccionan (luego de realizar una búsqueda en bases de datos tipo Scopus o PubMed) los estudios científicos donde se haya evaluado la eficacia de la prueba diagnóstica a analizar. Los estudios seleccionados deben aportar datos necesarios para estimar la eficacia de las pruebas diagnósticas en términos de sensibilidad y especificidad. En nuestro ejemplo se identificaron 88 estudios realizados para estimar la eficiencia de las técnicas de ELISA para la detección de leptospirosis en humanos.

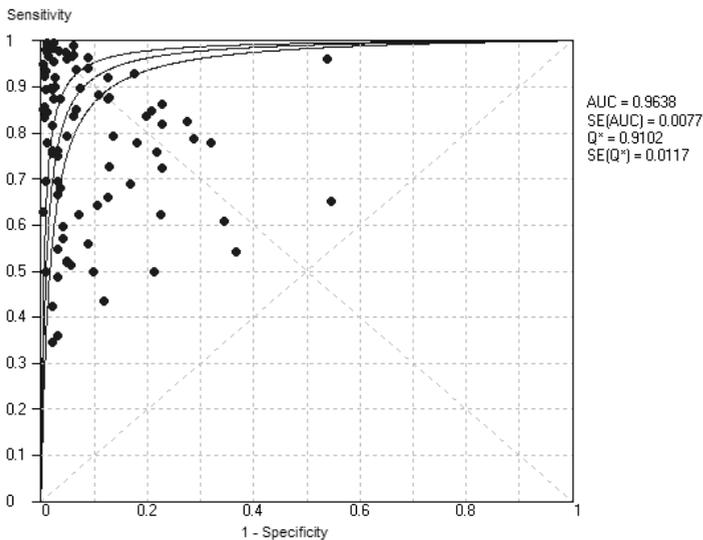
Posteriormente se calculan los estimadores globales de sensibilidad y especificidad que, para nuestro ejemplo, arrojaron una sensibilidad global de 0,779 (IC 95 % 0,770 – 0,789) y una especificidad global de 0,913 (IC 95 % 0,908 – 0,917). El grado de variabilidad entre los resultados de los estudios puede evaluarse gráficamente presentando la sensibilidad y especificidad de cada estudio en un *forest plot* (Figura 8). Alguna dispersión debería aparecer por el azar en la selección de las muestras en los estudios, pero otros factores la pueden aumentar.

Figura 8. *Forest plot* de la estimación de la sensibilidad y especificidad de las pruebas de ELISA para el diagnóstico de leptospirosis humana



No obstante, el mejor resumen de los resultados de los estudios es una curva ROC en lugar de sólo un índice global. La forma de la curva depende de las distribuciones de probabilidad subyacentes de los resultados de la prueba diagnóstica en los individuos con y sin la enfermedad. En nuestro ejemplo se obtuvo una curva ROC resumen con un área bajo la curva del 96,38 %, lo que indica una alta eficacia para estas pruebas (Figura 9). Cada punto en la curva ROC representa un estudio (aporta un par de valores de sensibilidad y especificidad).

Figura 9. Curva ROC resumen de las pruebas de ELISA para el diagnóstico de leptospirosis humana (Signorini et al., 2013)



A partir de los meta-análisis es posible estudiar la heterogeneidad en los estimados de sensibilidad y especificidad global e identificar las fuentes de dicha heterogeneidad a través de análisis estratificados. Mediante este análisis se calculan los estimados de sensibilidad y especificidad estratificando los estudios según características propias de los mismos que pudieran condicionar el desempeño de la prueba (ej.: fase de la enfermedad, tipo de Ig a detectar por el ELISA, edad de los individuos, entre otros factores).

Otra posibilidad para investigar fuentes de heterogeneidad es realizar una meta-regresión. En este modelo de regresión lineal se emplea el *Odds Ratio* Diagnóstico (el *ORD* expresa cuánto mayor es el *odds* de estar enfermo entre los individuos con resultado positivo que entre los que tienen resultado negativo. Es una medida simple del rendimiento diagnóstico de la prueba) como variable de respuesta. Las variables independientes (aquellas que se consideran pudieron afectar el desempeño de la prueba diagnóstica) son empleadas como co-variables con potencial de estar asociadas con la variabilidad en el ORD. Los resultados del modelo de meta-regresión

indican el cambio en el rendimiento diagnóstico de la prueba en estudio por aumento de una unidad en la co-variable correspondiente. En nuestro ejemplo, los estudios que incluyeron pacientes en fase convaleciente de la enfermedad mostraron un rendimiento significativamente mayor (1,84 veces) en comparación con aquellos estudios que emplearon pacientes en fase aguda o en los que no fue posible identificar la fase de la enfermedad que presentaban. Estudios basados en pruebas de ELISA que identificaban IgM produjeron un rendimiento que fue significativamente mayor (7,14 veces) que aquellos que identificaban IgG. Por otro lado, el tipo de antígeno usado en la prueba de ELISA no produjo un cambio significativo en el rendimiento de la prueba, lo que indica que el uso de antígenos a partir de células completas o recombinante/sintético no afectó significativamente la eficacia de la prueba.

Existen algunos programas, como el Meta-DiSc (Zamora et al., 2006), de libre distribución, que se usan para la realización de Meta-análisis de estudios de evaluación de pruebas diagnósticas. En veterinaria este tipo de estudios no es común ya que normalmente no existen muchos trabajos sobre validación de técnicas diagnósticas, por lo que el número de estudios a incluir generalmente es bajo y, por ende, se reduce la fortaleza de la herramienta.

Bibliografía

- Adler, H.E., Wiggins, A.D. (1973).** Interpretation of serologic tests for *Mycoplasma gallisepticum*. World. Poul. Sci. **29**: 345–349.
- Branscum, A.J., Gardner, I.A., Johnson, W.O. (2005).** Estimation of diagnostic-test sensitivity and specificity through Bayesian modeling. Preventive Veterinary Medicine **68**: 145–163.
- Briesofsky, J., Greiner, M. (1998).** Seroepidemiol Working Group, Inst. parasitol. & Trop. Vet. Med., Dept. Trop. Vet. Med. & Epidemiol. Freie Univ. Berlin, <http://city.vetmed.fu.berlin.de/~mgreiner/CMDT/cmdt.htm>
- Buck, A.A., Gart, G.G. (1966).** Comparison of a screening test and a reference test in epidemiologic studies. I. Indices of agreement and their relation to prevalence. Am. J. Epidemiol. **83**: 586–592.
- Cannon, R.M., Roe, R.T. (1982).** Livestock Disease Surveys: A Field Manual for Veterinarians. Aust. Bureau of Anim. Hlth., Dpt. Primary Ind., Aust. Gov. Pub. Serv., Canberra, 35 pp.
- Cohen, J. (1960).** A coefficient of agreement for nominal scales. Educ. Psychol. Meas. **20**: 37–46.
- Di Lorenzo, C., Olivera, M., Cabral, M., Argenio, L. (2009).** Validación, robustez y selectividad de la Reacción en Cadena de la Polimerasa Multiplex para la identificación de *Brucella canis*. Informe preliminar. 6tas Jornadas Internacionales de Veterinaria Práctica. Colegio de Veterinarios de la Provincia de Buenos Aires. Mar del Plata, 14 y 15 de agosto de 2009.
- Dohoo, I. R. (1981).** Effects of misclassification on statistical inferences in Epidemiology. Letter to the Editor. Am. J. Epidemiol. **118**: 485–486.
- Dohoo, I., Martin, W., Stryhn, H. (2003).** Veterinary Epidemiology Research. Ed. AVC Inc., Charlottetown, Prince Edward island (Canadá). Pp. 706. ISBN 0–919013–41–4.

- Donald, A.W. (1994).** Estimating the prevalence of mastitis and other organ-specific diseases in the presence of within animal disease correlation and diagnostic test correlation. *Prev. Vet. Med.* **20**: 113–133.
- Donald, A.W., Gardner, I.A., Wiggins, A.D. (1994).** Cut-off points for aggregate herd testing in the presence of disease clustering and correlation of test errors. *Prev. Vet. Med.* **19**: 167–187.
- Duffy, S. (1994).** Diagnosis at herd level. *Proc. Vllth Internat. Symposium Vet. Lab. Diag., Soc. Med. Vet. and Asoc. Arg. Vet. Lab. Diag., Buenos Aires*, p. 82.
- Elizalde, E.F., Signorini, M.L., Canavesio, V.R., Cuatrin, A., Tarabla, H.D., Calvino, L.F. (2009).** Medición de la conductividad eléctrica en leche como método de diagnóstico de mastitis subclínica bovina. *Revista FAVE-Ciencias Veterinarias*, **8(1)**:15–28.
- Fleiss, J.L. (1981).** *Statistical Methods for Rates and Proportions*, 2nd Ed., John Wiley & Sons Inc., New York, 321 pp.
- Galen, R. S. and Gambino, S. R. (1975).** Beyond Normality, the Predictive Value and Efficiency of Medical Diagnoses. John Wiley & Sons, New York, 237 pp.
- Gardner, I.A., Greiner, M. (1999).** Advanced Methods for test validation and interpretation in Veterinary Medicine. Freie Univ. Berlin, 74 pp.
- Gart, J.J., Buck, A.A. (1966).** Comparison of a screening test and a reference test in epidemiologic studies. II. A probabilistic model for the comparison of diagnostic tests. *Am. J. Epidemiol.* **83**: 593–602.
- Gray, M.D., Martin, S.W. (1980).** An evaluation of screening programs for the detection of brucellosis in dairy herds. *Can. J. Comp. Med.* **44**: 52–60.
- Greiner, M., Gardner, I.A. (2000).** Epidemiologic issues in the validation of veterinary diagnostic tests. *Preventive Veterinary Medicine*, **45**: 3–22.
- Hanley, J.A., McNeil, B.J. (1982).** The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. **143**: 29–36.
- Houe, H., Kjaer Ersboll, A., Toft, N., Agger, J.F. (2003).** Veterinary epidemiology: from hypothesis to conclusion. Ed. The Royal Veterinary and Agricultural University, Frederiksberg, Dinamarca, 340 pp.
- Hui, S.L., Walter, S.D. (1980).** Estimating the error rates of diagnostic tests. *Biometrics* **36**: 167–171.
- Knapp, R.G., Clinton Miller, M. (1992).** *Clinical Epidemiology and Biostatistics*. Natl. Med. Series, Williams & Wilkins, Harwal Publ. Co., Pennsylvania, 435 pp.
- Landis, J.R., Koch, G.G. (1977).** The measurement of observer agreement for categorical data. *Biometrics*. **33**: 159–174.
- Martin, S.W. (1976).** The evaluation of tests. *Can. J. Comp. Med.* **41**: 19–25.
- Martin, S.W. (1984).** Estimating disease prevalence and the interpretation of screening test results. *Prev. Vet. Med.*, **2**: 463–472.
- Martin, S.W., Shoukri, M., Thorburn, M. A. (1992).** Evaluating the health status of herds based on test applied to individuals. *Prev. Vet. Med.*, **14**: 33–43.
- Martin, S.W., Meek, A.H., Willeberg, P. (1987).** *Veterinary Epidemiology. Principles and Methods*. Iowa St. Univ. Press, Ames, Iowa, 343 pp.
- McNeil, B.J., Keeler, E., Adelstein, S.J. (1975).** Primer on certain elements of medical decision making. *N. Engl. J. Med.* **293**: 211–226.
- Metz, C. (1998).** *ROC Analysis*. Univ. Chicago, <http://www-radiology.uchicago.edu/cgi-bin/software.cgi>.
- Nielsen, K., Gall, P., Kelly, D., Henning, D., García, M. (1992).** Enzyme immunoassay. Application to diagnosis of bovine brucellosis. *Agric. Can., Ontario*, 203 pp.
- Pfeiffer, D. U. (1998).** *Herdacc*. Univ. Massey, New Zealand, <http://epiweb.massey.ac.nz/files/herdacc.exe>.
- Pouillot, R., Gerbier, G., Gardner, I.A. (2002).** "TAGS", a program for the evaluation of test accuracy in the absence of a gold standard. *Preventive Veterinary Medicine*, **53**: 67–81.
- Quade, D., Lachenbruch, P.A., Waley, F.S., McClish, D.K., Haley, R.W. (1980).** Effects of misclassification on statistical inferences in Epidemiology. *Am. J. Epidemiol.* **111**: 503–515.
- Riegelman R.K. & Hirsch R.P. Comunicación Biomédica.** Cómo estudiar un estudio y probar una prueba: lectura crítica de la literatura médica.
- Schoonjas, F., Zalata, A., Depuyt, c.E., Comhaise, F.H. (1995).** *MedCalc: a new computer*

program for medical statistics. *Comput. Methods & Programs* **48**: 257–262.

Seiler, R.J. (1979). The non-disease reactor: Considerations on the interpretation of screening test results. *Vet. Rec.* **105**: 226–228.

Signorini, M.L., Vanasco, N.B., Lottersberger, J., Tarabla, H.D. (2013). Enzyme-linked immunosorbent assay for diagnosis of Human Leptospirosis: A meta-analysis of the published literature. *Epidemiology and Infection*, **141**(1):22-32..

Snedecor, G.W., Cochran, W.G. (1980). *Statistical Methods*. Iowa St. Univ. Press, 7th Ed., 507 pp.

Tarabla, H.D. (1996). Importancia de la sensibilidad de la prueba diagnóstica y de la prevalencia de patógenos intramamarios en el tratamiento selectivo de la vaca seca. *Memorias Congr. Nac. Calidad de Leche y Mastitis*, Río Cuarto, Córdoba, pp. 21–23.

Tarabla, H.D. (2002). Validación de pruebas diagnósticas. *Rev. FAVE, Sec. Cien. Vet., UNL*, **1**: 37–42.

Thorner, R.M., Remein, Q.R. (1961). Principles and procedures in the evaluation of screening of disease. US Dept. Hlth. Educ. & Welfare, Public Hlth. Monograph N1 67, 24 pp.

Vanasco, N.B., Lottersberger, J., Schmeling, M.F., Gardner, I.A., Tarabla, H.D. (2007). Diagnóstico de leptospirosis: evaluación de un enzimoimmunoensayo en fase sólida en diferentes

etapas de la enfermedad. *Rev. Panam. Salud Pública* **21**(6): 388–395.

Vanasco NB, Lottersberger J, Guerrero SA, Chiani Y, Schmeling MF, Tarabla H D. (2011). Evaluation of ELISA antigens for leptospiral serodiagnosis in different stages of the disease. VII Reunión de la Sociedad Internacional de Leptospirosis, Merida, Yucatan, Mexico, 19–22 de Septiembre, 2011. Oral presentation.

Wiggins, A.D., Franti, C.E. (1980). An inverse sampling scheme using blocks. *Math. Biosci.* **49**: 139–153.

Yamamoto, R. (1975). Characteristics an evaluation of screening tests. *Mycoplasmosis Workshop Am. Assoc. Avian Pathol. 112th Annu. Mtg. Am. Vet. Med. Assoc., Anaheim, California. Univ. California Davis, EPM Dpt., 5 pp.*

Yamamoto, R. (1983). Factors affecting sensitivity and specificity of serologic tests: Some practical considerations and general thoughts. *Univ. California Davis, EPM 216 Notes*, 5 pp.

Yerushalmy, J. (1947). Statistical problems in assessing methods of medical diagnosis, with special reference to X-ray techniques. *Pub. Hlth. Rep.* **62**: 1432–1449.

Zamora, J., Abaira, V., Muriel, A., Khan, K., Coomarasamy, A. (2006). Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC Medical Research Method.* **6**: 31–42.

Capítulo 7

Prevalencia real y prevalencia aparente

1. Prevalencia e incidencia

Prevalencia es una cifra relativa que relaciona el número de animales enfermos con una población susceptible, un espacio y un tiempo determinado. En otras palabras, la prevalencia denota la probabilidad de que un animal esté enfermo dentro de una población [P(E+)]. No es una tasa verdadera (Eland Johnson, 1975) sino una proporción utilizada para cuantificar la frecuencia de enfermedad en una población en un momento determinado en el tiempo (prevalencia puntual) y se define como:

Prevalencia:

$$\frac{\text{Total de enfermos en la unidad de tiempo}}{\text{Población expuesta al riesgo}} \times 10^n$$

Incidencia, por otra parte, es el número relativo de nuevos casos de la enfermedad en una población, un espacio y un tiempo determinados, definida como:

Tasa de incidencia:

$$\frac{\text{Total de nuevos casos en la unidad de tiempo}}{\text{Población expuesta al riesgo}} \times 10^n$$

No se debe confundir prevalencia o tasa de incidencia con la proporción de enfermos de una afección en particular que ingresa en una clínica o centro de diagnóstico. Estas cifras relativas, donde el denominador es el total de pacientes o muestras ingresadas se denominan tasas proporcionales y no necesariamente reflejan la prevalencia o incidencia de la afección en la población (Tarabla, 1987).

En los estudios poblacionales, las pruebas para detectar enfermedad son utilizadas mayormente para estimar la prevalencia de la enfermedad de interés. En estos casos se usufructúa la habilidad de la prueba para detectar anticuerpos persistentes, lo cual permite detectar infecciones u otros eventos producidos años anteriores. Sin embargo, en algunas situaciones se desea conocer aunque sea en forma aproximada la incidencia de una enfermedad. Para ello se usan pruebas que detectan infecciones recientes (ej.: inmunofluorescencia directa en rinotraqueítis infecciosa bovina) o se divide la población a muestrear en estratos etarios (ej.: bovinos menores de dos años para detectar la circulación reciente de virus aftoso en un área).

2. Prevalencia real

Se denomina prevalencia real a la proporción de animales enfermos existentes en el estado natural, que es independiente y anterior a la ejecución de la prueba (Cuadro 1).

Cuadro 1. Prevalencia real

Condición	Enfermos	Sanos	Total
Test positivos	PV	FP	PV + FP
Test negativos	FN	NV	NV + FN
Total	PV + FN	NV + FP	PV + FP + NV + FN = n

$$\text{Prevalencia real} = \text{PR} = \frac{\text{PV} + \text{FN}}{n} \times 100$$

3. Prevalencia aparente

Se denomina prevalencia aparente a la proporción de animales positivos a la prueba diagnóstica (Cuadro 2).

Cuadro 2. Prevalencia aparente

Condición	Enfermos	Sanos	Total
Test positivos	PV	FP	PV + FP
Test negativos	FN	NV	NV + FN
Total	PV + FN	NV + FP	PV + FP + NV + FN = n

$$\text{Prevalencia aparente} = \text{PA} = \frac{\text{PV} + \text{FP}}{n} \times 100$$

La proporción de animales positivos a una prueba diagnóstica es una estimación sesgada de la verdadera prevalencia de la enfermedad. La magnitud del sesgo dependerá de las características de la prueba y de la prevalencia real de la enfermedad (Rogan y Gladen, 1978).

4. Prevalencia real y prevalencia aparente. Un ejemplo

Supongamos que existe un país con una población bovina de 100 000 animales en la que hay prevalencia real de tuberculosis del 1 %. En esta población efectuamos una prueba tuberculínica simple en tabla del cuello con PPD que puede llegar a tener una sensibilidad y especificidad máximas del 81,8 y 96,3 %, respectivamente (Francis et al., 1978) (Cuadro 3).

Cuadro 3. Clasificación de una población bovina con respecto a su estado de salud en tuberculosis (prevalencia real del 1 %) y los resultados de la prueba tuberculínica intradérmica simple

Condición	Tuberculosos	No tuberculosos	Total
Test positivos	818	3663	4481
Test negativos	182	95 337	95 519
Total	1000	99 000	100 000

Como es dable observar, en esta población con una prevalencia real del 1 % sobreestimaríamos la misma en más de 4 veces, dado que la prevalencia aparente es del 4,48 %. Si analizamos esto en el contexto de un programa de erradicación de la enfermedad, significa que estaremos subestimando groseramente los progresos del programa (Tarabla, 1995). En este caso, la probabilidad de que un animal que nos de positivo a la prueba esté en realidad sano es muy alta:

$$\frac{3363}{4481} \times 100 = 81,7\%$$

En el Capítulo 8 se desarrollará este tema más profundamente, mientras que en el Capítulo 11 se discutirá el cálculo de la prevalencia real a partir de la aparente.

Referencias

- Elandt-Johnson, R.C. (1975).** Definition of rates: Some remarks on their use and misuse. *Am. J. Epidemiol.* **102**: 267–271.
- Francis, J., Seiler, R.J., Wilkie, I.W., O’Bye, D., Lumsden, M.J., and Frost, A.J. (1978).** The sensitivity and specificity of various tuberculin tests using bovine PPD and other tuberculins. *Vet. Rec.* **103**: 420–435.
- Rogan, W.J. and Gladen, B. (1978).** Estimating prevalence from the results of a screening test. *Am. J. Epidemiol.* **107**: 71–76.
- Tarabla, H.D. (1987).** Cuantificación de la ocurrencia de enfermedades. *Therios.* **9**: 25–31.
- Tarabla, H.D. (1995).** Efecto del poder discriminatorio de la prueba diagnóstica sobre la estimación de la prevalencia de la enfermedad. Resúmenes **Congr. Arg. y 11 Congr. Latinoam.** Zoonosis, Buenos Aires, Argentina, p. 209.

Capítulo 8

Valor predictivo

1. Valor predictivo de la prueba positiva

El valor predictivo de la prueba positiva o valor predictivo positivo (VP+) es la probabilidad de que un animal positivo a la prueba esté realmente enfermo ($P(E+|T+)$) (Cuadro 1).

Cuadro 1. Valor predictivo de una prueba diagnóstica positiva

Condición	Enfermos	Sanos	Total
Test positivos	PV	FP	PV + FP
Test negativos	FN	NV	NV + FN
Total	PV + FN	NV + FP	PV + FP + NV + FN = n

$$\text{Valor predictivo positivo} = VP+ = \frac{PV}{PV + FP} \times 100$$

2. Valor predictivo de la prueba negativa

El valor predictivo de la prueba negativa o valor predictivo negativo (VP-) es la probabilidad de que un animal negativo a la prueba esté realmente sano ($P(E-|T-)$) (Cuadro 2).

Cuadro 2. Valor predictivo de una prueba diagnóstica negativa

Condición	Enfermos	Sanos	Total
Test positivos	PV	FP	PV + FP
Test negativos	FN	NV	NV + FN
Total	PV + FN	NV + FP	PV + FP + NV + FN = n

$$\text{Valor predictivo negativo} = VP- = \frac{NV}{NV + FN} \times 100$$

3. Determinación del valor predictivo de la prueba

3.1. Valor predictivo positivo

3.1.1. Teorema de Bayes

En la práctica diaria, un clínico puede estimar conforme a su experiencia cuál es la probabilidad de que en una enfermedad dada (ej.: brucelosis) un animal presente un síntoma (ej.: aborto) o positividad a una prueba diagnóstica determinada (ej.: aglutinación). En otras palabras, la probabilidad de que el animal aborte siendo bruceloso $Pr(\text{Síntoma}|\text{Enfermedad})$ o la probabilidad de que sea positiva la reacción de aglutinación estando enfermo de brucelosis $Pr(T+|E+)$. Este esquema es coherente con la curricula clínica de la mayoría de las universidades y con el formato de los libros de texto, que describen los signos probables cuando una determinada enfermedad está presente. Sin embargo, el clínico se ve enfrentado diariamente con la tarea opuesta. En realidad, él debe decidir cuál es el diagnóstico más probable ante una serie de signos, síntomas y/o hallazgos de laboratorio. Es decir, lo que realmente interesa es la probabilidad de que un animal padezca brucelosis ante un caso de aborto $P(\text{Enfermedad}|\text{Síntoma})$ o ante una reacción serológica positiva $P(E+|T+)$ (Edwards et al., 1997). La fórmula para hallar esta probabilidad es conocida como Regla, Fórmula o Teorema de Bayes (Bayes, 1763).

$$Pr(E/+)=\frac{Pr(E+)\times Pr(T+|E+)}{Pr(E+)\times Pr(T+|E+)+Pr(E-)\times Pr(T+|E-)}$$

Donde:

$Pr(E+)$ = Probabilidad de enfermedad (prevalencia de la enfermedad en la población)

$Pr(E-)$ = Probabilidad de sano

$Pr(T+|E+)$ = Probabilidad de que estando enfermo sea positivo a la prueba diagnóstica

$Pr(T+|E-)$ = Probabilidad de que estando sano sea positivo a la prueba diagnóstica

Podemos reemplazar estas probabilidades por elementos ya presentados con anterioridad.

$Pr(E+|T+)$ = $VP+$ = Valor predictivo positivo

$Pr(E+)$ = PR = Prevalencia real de la enfermedad

$Pr(E-)$ = $1 - PR$ = Prevalencia de sanos

$Pr(T+|E+)$ = $Sens$ = Sensibilidad de la prueba diagnóstica

$Pr(T+|E-)$ = $1 - Esp$ = Proporción de positivos sanos

De esta forma, el valor predictivo positivo quedaría:

$$\text{Valor predictivo positivo} = \frac{\text{Prevalencia} \times \text{Sens}}{\text{Prevalencia} \times \text{Sens} + (1 - \text{Prevalencia}) \times (1 - \text{Esp})}$$

Supongamos que contamos con una población (n) de 10 000 animales en la cual está presente una enfermedad que tiene una prevalencia real del 10 % y para diagnosticarla se emplea una prueba que tiene una sensibilidad y una especificidad equivalentes al 95 %.

$$\begin{aligned} \text{VP+} &= \frac{0,10 \times 0,95}{0,10 \times 0,95 + (1 - 0,10) \times (1 - 0,95)} = \\ &= \frac{0,095}{0,14} = 0,679 \text{ o } 67,9 \% \end{aligned}$$

Esto significa que, ante un resultado positivo a la prueba diagnóstica, la probabilidad de que ese individuo esté realmente enfermo es del 67,9 %.

3.1.2. Cuadro de cálculo del valor predictivo

El Teorema de Bayes puede resultar engorroso e innecesario puesto que simplemente extrapola en forma horizontal la información proporcionada por la tabla de 2 x 2 presentada con anterioridad en el Capítulo 4. El mismo cálculo efectuado mediante el Teorema de Bayes puede ser desarrollado en forma algo intuitiva utilizando para ello el Cuadro 2 del Capítulo 4 (Slensing y Gardner, 1997). Como primer paso se anota el n (Cuadro 3).

Cuadro 3. Población en estudio

Condición	Enfermos	Sanos	Total
Test positivos			
Test negativos			
Total			10 000

Luego se calcula el número de animales enfermos (PV + FN) de acuerdo a la prevalencia real que estipulamos era del 10 %, y por sustracción [n - (PV + FN)] la cantidad de animales sanos (NV + FP) (Cuadro 8.4).

Cuadro 4. Número de enfermos y sanos

Condición	Enfermos	Sanos	Total
Test positivos			
Test negativos			
Total	1000	9000	10 000

Si la prueba tiene una sensibilidad del 95 %, significa que el 95 % de los enfermos va a estar correctamente clasificado como tal (PV) y el restante 5 % son animales enfermos no detectados como positivos por la prueba (FN) (Cuadro 5).

Cuadro 5. Número de positivos verdaderos, falsos negativos, enfermos y sanos

Condición	Enfermos	Sanos	Total
Test positivos	950		
Test negativos	50		
Total	1000	9000	10 000

De la misma manera, si la prueba tiene una especificidad del 95 %, significa que el 95 % de los sanos va a estar correctamente clasificado como tal (NV) y que el 5 % son animales sanos no detectados como tales por la prueba (FP) (Cuadro 6).

Cuadro 6. Número de negativos verdaderos, falsos positivos, positivos verdaderos, falsos negativos, enfermos y sanos

Condición	Enfermos	Sanos	Total
Test positivos	950	450	
Test negativos	50	8550	
Total	1000	9000	10 000

Por último, la suma PV + FP nos dará el total de animales positivos a la prueba y la suma NV + FN la cantidad de animales negativos a la misma (Cuadro 7).

Cuadro 7. Cuadro de cálculo del valor predictivo positivo

Condición	Enfermos	Sanos	Total
Test positivos	950	450	1400
Test negativos	50	8550	8600
Total	1000	9000	10 000

Siendo el valor predictivo positivo equivalente a:

$$VP+ = \frac{PV}{PV + FP} \times 100$$

De acuerdo con los valores del Cuadro 7 tendremos:

$$VP+ = \frac{950}{950 + 450} \times 100 = 67,9 \%$$

que es exactamente lo calculado en el punto **3.1.1.** según el Teorema de Bayes. Nótese además que en este ejemplo la prevalencia aparente es del 14 %, mientras que la real era sólo del 10 %.

3.2. Valor predictivo negativo

El valor predictivo de la prueba negativa ($VP-$) es de particular interés cuando se quieren introducir animales a una zona libre o con programas de control de la enfermedad de interés. Se define como la probabilidad de que un individuo negativo a la prueba diagnóstica esté realmente sano (Bayes, 1763).

$$VP- = \frac{(1 - PR) \times Esp.}{(1 - PR) \times Esp + PR \times (1 - Sens)}$$

Siguiendo el ejemplo hipotético del punto anterior, tendremos:

$$VP- = \frac{(1 - 0,10) \times 0,95}{(1 - 0,10) \times 0,95 + 0,10 \times (1 - 0,95)} =$$

$$VP- = \frac{0,855}{0,86} = 0,994 \text{ o } 99,4 \%$$

Esto significa que un individuo negativo a la prueba diagnóstica tiene un 99,4 % de posibilidades de estar realmente sano.

Si queremos utilizar el método de cálculo del cuadro de valor predictivo, tal como se hizo para el valor predictivo positivo y repitiendo los valores del Cuadro 7, tendremos (Cuadro 8):

Cuadro 8. Cuadro de cálculo del valor predictivo negativo

Condición	Enfermos	Sanos	Total
Test positivos	950	450	1400
Test negativos	50	8550	8600
Total	1000	9000	10 000

Siendo el valor predictivo negativo equivalente a:

$$VP_- = \frac{NV}{NV + FN} \times 100$$

$$VP_- = \frac{8550}{50 + 8550} \times 100 = 99,4 \%$$

3.3. Valores predictivos. Un ejemplo

Tomemos ahora como ejemplo la determinación de la presencia de residuos de antibióticos en leche de vaca que tiene, en promedio, una sensibilidad y especificidad aproximada del 91,0 % y 81,0 %, respectivamente (Slennig y Gardner, 1997). Siguiendo con la misma mecánica comenzada en el «Cuadro de cálculo del valor predictivo», si aplicamos esta prueba en una población de 10 000 animales donde la prevalencia de residuos de antibióticos en leche es del 2 %, obtendremos (Cuadro 9):

Cuadro 9. Clasificación de una población bovina con respecto a la presencia de residuos antibióticos en leche (prevalencia real 2 %) y los resultados de una prueba para detección de inhibidores

Condición	Con residuos	Sin residuos	Total
Test positivos	182	1862	2044
Test negativos	18	7938	7956
Total	200	9800	10 000

$$\text{Sens} = \frac{182}{200} \times 100 = 91,0 \%$$

$$\text{Esp} = \frac{7938}{9800} \times 100 = 81,0 \%$$

$$VP_+ = \text{Pr}(E+ | T+) = \frac{182}{2044} \times 100 = 8,9 \%$$

$$VP- = \Pr (E- | T-) = \frac{7938}{7956} \times 100 = 99,8 \%$$

Esto significa que, si una muestra de leche es seleccionada al azar de una población en donde el 2 % de las leches contiene residuos de antibióticos y es examinada con una prueba para detección de inhibidores, la probabilidad de que no contenga residuos habiendo sido negativa a la prueba es cercana al 100 %. De igual manera, la probabilidad de que, habiendo dado positiva a la prueba, la muestra contenga residuos, es de sólo el 8,9 %.

4. Determinación de la tasa de falsos positivos

Aunque algunos autores la consideran como complemento de la especificidad, en este texto la tasa de falsos positivos es la proporción de animales sanos dentro del total de resultados positivos a la prueba y puede ser calculada como:

$$\text{Tasa de FP} = \frac{FP}{FP + PV} \cdot 10^n$$

Siguiendo el ejemplo del Cuadro 7:

$$\text{Tasa de FP} = \frac{450}{450 + 950} = 32,14 \%$$

Sin necesidad de construir la tabla de cálculo del valor predictivo, esta tasa se puede determinar siguiendo los principios del Teorema de Bayes. (Bayes, 1763)

$$\text{Tasa de FP} = \frac{(1 - PR) \times (1 - \text{Esp.})}{\text{Sens} \times PR + (1 - PR) \times (1 - \text{Esp.})}$$

En el ejemplo del Cuadro 9, la tasa de falsos positivos es muy alta (91,1 %). El costo de los FP sería, en este caso, afrontado por los propietarios de los establecimientos productores de leche cuya producción sería penalizada.

5. Determinación de la tasa de falsos negativos

Aunque en algunos casos es considerada como el valor complementario de la sensibilidad, en este texto la tasa de falsos negativos es la proporción de animales enfermos dentro del total de resultados negativos a la prueba:

$$\text{Tasa de FN} = \frac{\text{FN}}{\text{FN} + \text{NV}} \cdot 10^n$$

Siguiendo el ejemplo del Cuadro 8.7:

$$\text{Tasa de FN} = \frac{50}{50 + 8550} = 0,58 \%$$

Tomando como base el Teorema de Bayes, para un individuo esta probabilidad se calcula de la siguiente manera (McClish y Quade, 1985):

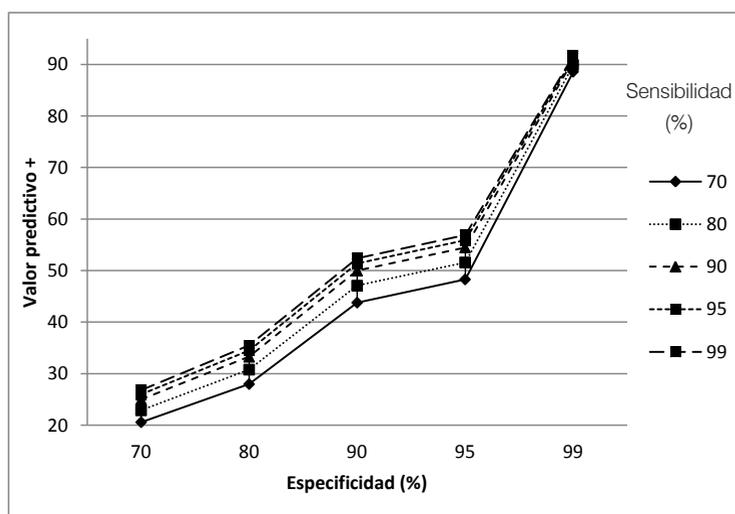
$$\text{Tasa de FN} = \frac{\text{PR} \times (1 - \text{Sens})}{\text{PR} \times (1 - \text{Sens}) + (1 - \text{PR}) \times \text{Esp}} =$$

En el ejemplo del Cuadro 9. la tasa de falsos negativos es del 0,2 %. El costo de los FN sería afrontado por la industria láctea, cuyos procesos podrían ser afectados por la presencia de antibióticos en leche.

6. Efecto de la sensibilidad y especificidad de la prueba sobre su valor predictivo

Dos factores que influyen marcadamente en los valores predictivos son la sensibilidad y la especificidad de la prueba diagnóstica. El VP+ se ve más afectado por cambios en la especificidad de la prueba que por variaciones en la sensibilidad (Figura 1).

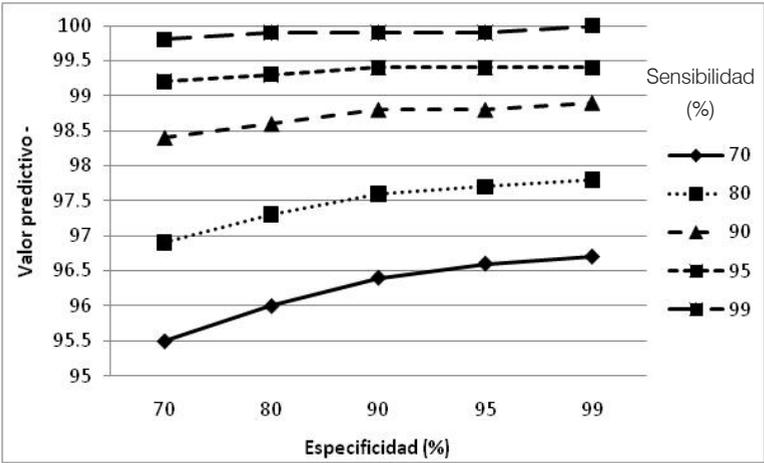
Figura 1. Valor predictivo de la prueba positiva en diversas combinaciones de sensibilidad y especificidad cuando la prevalencia real de la enfermedad es del 10 %



Como se observa en la Figura anterior, al aumentar la especificidad se incrementa notablemente el VP+, mientras que esos cambios son menos dramáticos cuando varía la sensibilidad de la prueba diagnóstica.

Como contrapartida, el VP- se ve más afectado por variaciones en la sensibilidad de la prueba. Al aumentar la sensibilidad, el VP- aumenta en una proporción mayor (aunque menos notoria que en el caso de las variaciones del VP+ ante cambios en la especificidad) que cuando se producen idénticas variaciones en la especificidad (Figura 2).

Figura 2. Valor predictivo de la prueba negativa en diversas combinaciones de sensibilidad y especificidad cuando la prevalencia real de la enfermedad es del 10 %



Resumiendo, a medida que aumenta la sensibilidad de la prueba, aumenta el VP- y disminuyen los diagnósticos falsos negativos. Por otra parte, al aumentar la especificidad de la prueba, aumenta el VP+ y disminuyen los diagnósticos falsos positivos.

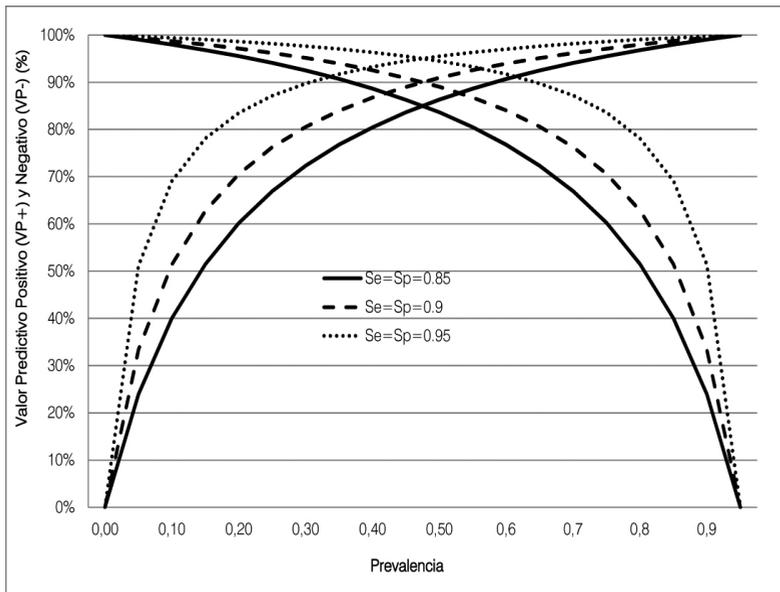
7. Efecto de la prevalencia real de la enfermedad sobre el valor predictivo de la prueba

A medida que disminuye la prevalencia real de la enfermedad disminuye el valor predictivo positivo. Este efecto es tangible en los programas de control de enfermedades exitosos. A medida que el programa progresa y va disminuyendo la frecuencia de la enfermedad en la población, ésta se va transformando paulatinamente en una comunidad sana. En las fases avanzadas del programa, y asumiendo que la sensibilidad y la especificidad de la prueba permanecen constantes, la disminución del VP+ hace que aumente la cantidad de diagnósticos falsos positivos. Este aumento probablemente sea menos espectacular que lo aparentado, dado que lo más afectado es la proporción

de diagnósticos ($1 - VP+$) (Yamamoto, 1975:5), como se verá en el punto siguiente, «Efecto de la prevalencia real sobre los valores predictivos, los falsos positivos y los falsos negativos».

Por otra parte, a medida que baja la prevalencia real de la enfermedad, sube el valor predictivo negativo y baja la cantidad de falsos negativos. A pesar de esta disminución, estos diagnósticos negativos en animales enfermos también pueden llamar la atención durante las últimas etapas de un plan de control de una enfermedad (Yamamoto, 1975:5), como se ejemplificará en el punto 9. En tanto, en forma inversa, cuando aumenta la prevalencia real de la enfermedad, disminuye el valor predictivo negativo y aumenta la cantidad de falsos negativos. Este efecto sobre el $VP-$ sólo se hace notable en casos de prevalencias reales muy elevadas (Francis et al., 1978). Cuando la prevalencia es del 50 %, el $VP+$ es igual a la sensibilidad y el $VP-$ es igual a la especificidad (Figura 3).

Figura 3. Valores predictivos positivos y negativos en distintas prevalencias reales cuando la sensibilidad y especificidad de la prueba son del 85, 90 y 95 %.



Lógicamente, el $VP+$ aumenta en forma significativa en los grupos de alto riesgo, donde es dable esperar una mayor prevalencia de la afección de interés.

8. Efecto de la prevalencia real sobre los valores predictivos, los falsos positivos y los falsos negativos. Ejemplos

8.1. Sobre el valor predictivo de la prueba positiva

Tomando como ejemplo una población de 1000 caballos con una prevalencia real de Anemia Infecciosa Equina (AIE) del 30 %, en el cual se aplica la prueba de Inmunodifusión en Gel (IDG) y asumiendo que la misma tiene una sensibilidad y especificidad cercanas al 95,0 % (Coggins et al., 1972), se obtendría (Cuadro 10).

Cuadro 10. Clasificación de una población equina con respecto a su estado de salud en Anemia Infecciosa Equina (AIE, prevalencia real del 30 %) y los resultados de la prueba de Inmunodifusión en Gel (IDG)

Condición	Con AIE	Sin AIE	Total
Test positivos	285	35	320
Test negativos	15	665	680
Total	300	700	1000

En este caso, la probabilidad de estar realmente enfermo de AIE siendo positivo a la IDG es de casi el 90 %.

$$VP+ = \Pr (E+ | T+) = \frac{285}{320} \times 100 = 89,1 \%$$

Supongamos ahora que en la mencionada población equina se aplica un programa de control de AIE con buenos resultados, de manera tal que se reduce la prevalencia real de la enfermedad al 2 %. Si se sigue aplicando la IDG para su diagnóstico, obtendríamos (Cuadro 11):

Cuadro 11. Clasificación de una población equina con respecto a su estado de salud en Anemia Infecciosa Equina (AIE, prevalencia real del 2 %) y los resultados de la prueba de Inmunodifusión en Gel (IDG)

Condición	Con AIE	Sin AIE	Total
Test positivos	19	49	68
Test negativos	1	931	932
Total	20	980	1000

Por lo tanto, si un animal es seleccionado al azar en una población en donde el 2 % de sus integrantes tiene AIE y es examinado con la IDG, la probabilidad de estar realmente enfermo siendo positivo es de sólo el 27,9 % *versus* el 89,1 % calculado antes de la implementación del programa de control.

$$VP+ = \Pr (E+ | T+) = \frac{19}{68} \times 100 = 27,9 \%$$

De esta manera, la posibilidad de eliminar un animal verdaderamente sano que resultó positivo a la IDG aumentó en forma proporcional a la disminución del valor predictivo de la prueba positiva.

8.2. Sobre los falsos positivos

En el ejemplo anterior, nótese que si bien aumentó el número de falsos positivos (de 35 animales se pasó a 49), el aumento en la *proporción* de diagnósticos fue mucho más notorio. Cuando la prevalencia real era del 30 %, la proporción de diagnósticos falsos positivos era:

$$\frac{35}{320} \times 100 = 10,9 \%$$

Sin embargo, al caer la prevalencia al 2 % dicha proporción subió a más del 70 %

$$\frac{49}{68} \times 100 = 72,1 \%$$

8.3. Sobre el valor predictivo de la prueba negativa

Tomemos ahora la situación inversa. En un contexto en el cual la prevalencia de enfermedad aumenta en una región tendríamos que, cuando la prevalencia es del 2 %, el VP- es igual a:

$$VP- = \Pr (E-/T-) = \frac{931}{932} \times 100 = 99,9 \%$$

Esto significa que, si la AIE aumentara su prevalencia del 2 al 30 % y seleccionáramos al azar un animal negativo a la prueba de Coggins, la probabilidad de que un individuo esté verdaderamente sano descendería del 99,9 al 97,8 %. De igual manera aumentarían los animales falsos negativos de 1 a 15.

8.4. Sobre los falsos negativos

Aunque resulte paradójico, en los pasos finales de un plan de control de una enfermedad, cuando la prevalencia real de la enfermedad es baja y, por ende, la proporción de verdaderos positivos es comparativamente pequeña, uno suele prestar más atención a aquellos animales enfermos que dan negativos a la prueba (falsos negativos).

El problema de los falsos negativos ya existía al comienzo de la campaña (15 animales cuando la prevalencia de AIE era del 30 %), con una proporción de falsos negativos:

$$\frac{15}{680} \times 100 = 2,2 \%$$

pero dada la alta carga de verdaderos positivos que existía, a nadie le importaba la existencia de animales enfermos no detectados por la prueba.

No obstante, al final del programa, cuando se intente la erradicación de la enfermedad, uno puede prestar mayor atención a los diagnósticos falsos negativos, a pesar de que en realidad haya menos reaccionantes de este tipo. Por ejemplo, al bajar la prevalencia de AIE al 2 % por efecto del programa implementado, sólo hubo 1 diagnóstico falso negativo, lo que representó una proporción sobre el total de negativos de solamente

$$\frac{1}{932} \times 100 = 0,1 \%$$

Como ejemplos de situaciones en las cuales cobra importancia la relación directamente proporcional que existe entre la prevalencia real de la enfermedad y los falsos negativos tenemos la inspección veterinaria en frigoríficos y el tratamiento de vaca seca en razas lecheras.

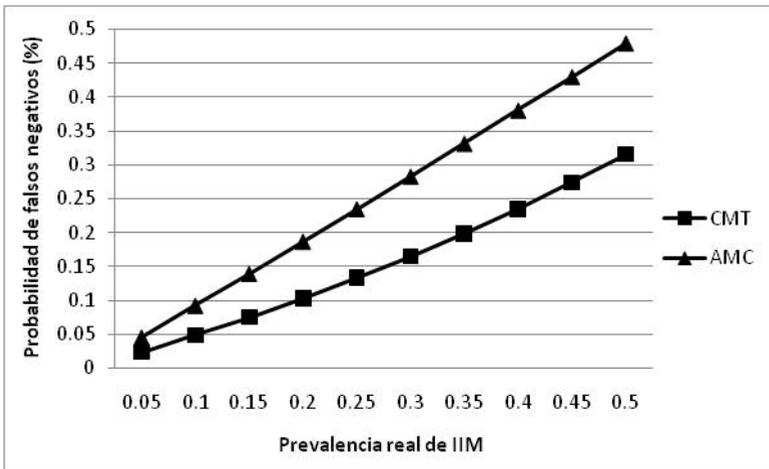
Cuando el sistema tradicional de inspección de carnes fue desarrollado al final del siglo pasado, la tuberculosis y la hidatidosis significaban grandes riesgos para la salud pública. Al ser muy elevada la prevalencia de estas enfermedades, la probabilidad de clasificar como sanos animales que en realidad estaban enfermos era también elevada (Greiner et al., 1981). A medida que estas afecciones se fueron controlando en la población, la probabilidad de diagnósticos falsos negativos, de vital importancia en este caso, fue también disminuyendo. Como contrapartida, se fue pasando gradualmente a la situación descrita en el punto "Sobre los falsos positivos", en la cual en bajas prevalencias reales disminuyó la habilidad de los inspectores para detectar animales verdaderamente enfermos.

En el segundo caso, la llamada terapia de vaca seca constituye uno de los elementos más importantes para el control de la mastitis bovina y consiste en la infusión por vía intramamaria de antibióticos de liberación lenta para curar las infecciones intramamarias presentes y prevenir las que se puedan producir durante el período no lactante. Una de las formas de aplicarla es la llamada terapia selectiva, teniendo como criterios de selección de los animales infectados los resultados de pruebas indirectas como el California Mastitis Test (CMT) o los Antecedentes de Mastitis Clínicas (AMC) (Tarabla et al., 1994). Estos resultados son expresados, en esos casos, en forma cualitativa como positivos o negativos al CMT, o bien como tiene o no tiene AMC.

Dado que la probabilidad de FN está influenciada especialmente por variaciones en la sensibilidad de la prueba, mientras que la especificidad tiene en estos casos un rol secundario (Tarabla, 1995), se tomó como base lo publicado por Rindsig *et al* (1979), donde CMT y AMC tenían idéntica especificidad (CMT: Sens: 75,0 %, Esp: 54,2 %; AMC: Sens: 50,0 %, Esp 54,2 %). Nótese que el poder discriminatorio de los AMC no es mejor que el que tendría el «tirar la moneda». Los inconvenientes que complican la estimación de la sensibilidad y especificidad en las enfermedades multicausales que producen respuestas medibles de distinta magnitud de acuerdo con el patógeno actuante ya fueron discutidos con anterioridad. Por razones prácticas se asumió que las pruebas tenían sensibilidad y especificidad estables en todas las IIM, pero se debe tener en claro que las cifras resultantes en los cálculos de los valores predictivos y proporciones de falsos negativos no son válidos para la extrapolación de su valor absoluto sino sólo como indicativos de las tendencias que asumirán al variar la prevalencia real de infecciones intramamarias y los tamaños de los rodeos.

A medida que aumenta la prevalencia real de la enfermedad aumenta la cantidad de FN, efecto que se hace notable en pruebas con baja sensibilidad (Figura 4).

Figura 4. Asociación entre la prevalencia real de Infecciones Intramamarias (IIM) y la proporción de falsos negativos a la prueba de mastitis de California (CMT) o los Antecedentes de Mastitis Clínicas (AMC)



Si aplicamos el CMT como prueba tamiz en un rodeo de 100 animales en ordeño con una prevalencia real del 50 %, existe la probabilidad de que aproximadamente un animal de cada 100 sea un diagnóstico FN (Cuadro 12).

Cuadro 12. Clasificación de una población bovina con respecto a su estado de salud en IIM (prevalencia real del 5 %) y los resultados de la prueba del CMT

Condición	Con IIM	Sin IIM	Total
CMT positivos	4	44	48
CMT negativos	1	51	52
Total	5	95	100

Sin embargo, si en el mismo rodeo la prevalencia de IIM fuera del 30 %, luego de efectuado el CMT prácticamente 8 animales de cada 100 pueden ser diagnósticos FN (Cuadro 16).

Cuadro 13. Clasificación de una población bovina con respecto a su estado de salud en IIM (prevalencia real del 30 %) y los resultados de la prueba del CMT

Condición	Con IIM	Sin IIM	Total
CMT positivos	22	32	54
CMT negativos	8	38	64
Total	30	70	100

Si, en lugar del CMT, aplicáramos los AMC como prueba tamiz en un rodeo de 100 animales en ordeño con una prevalencia real de IIM del 5 %, existe la probabilidad de que aproximadamente tres animales de cada 100 sean diagnósticos FN. Si la prevalencia de IIM fuera del 30 %, habría 15 animales FN. Por otra parte, es claro que a medida que aumenta la prevalencia real de IIM aumenta la cantidad de FN, lo que torna poco aconsejable el uso del tratamiento selectivo en rodeos con alta prevalencia de IIM (Rindsig et al., 1979).

10. Efecto del tamaño de la muestra sobre la proporción de falsos negativos

En diversas oportunidades se requiere que los animales sean negativos a una o más pruebas diagnósticas para permitir su entrada en áreas o países protegidos con barreras sanitarias (ej.: exportación de vaquillonas lecheras de Argentina a Brasil negativas a brucelosis y leucosis). En estos casos, para el exportador lo más importante es evitar los reactores falsos positivos que entorpezcan la certificación de su estado de libre de la enfermedad (Lloyd-Webb et al., 1995). Sin embargo, si el área de origen no es efectivamente libre, para el importador siempre existe el riesgo de introducir animales enfermos que den negativos a las pruebas diagnósticas (ej.: falsos negativos).

Retomemos ahora el ejemplo desarrollado en el punto «Prevalencia real y prevalencia aparente. Un ejemplo». En un país con una prevalencia de tuberculosis bovina del 1 % efectuamos una prueba tuberculínica simple en tabla del cuello con PPD

(Sens= 91 %, Esp = 89 %) (Francis et al., 1978). En este caso, utilizando la fórmula del punto «Determinación de la tasa de falsos negativos», la tasa de falsos negativos, es decir la probabilidad de que un reaccionante negativo esté en realidad enfermo, Pr(E+/T-), será igual a:

$$\begin{aligned} \text{Tasa de FN} &= \frac{0,01 \times (1 - 0,91)}{0,01 \times (1 - 0,91) + (1 - 0,01) \times 0,89} = \\ &= \frac{0,0009}{0,081} = 0,0111 \text{ o } 1,1 \% \end{aligned}$$

Esta probabilidad está marcadamente influenciada por la sensibilidad de la prueba y nos indica que podemos esperar que sólo un 1,11 % de animales negativos a la tuberculina estén verdaderamente enfermos.

Como vimos en el punto «Efecto de la prevalencia real de la enfermedad sobre el valor predictivo de la prueba», el valor predictivo negativo varía muy poco y se mantiene en niveles muy altos cuando la prevalencia de la enfermedad es baja. Por ello, la probabilidad de falsos negativos se mantiene en niveles muy bajos aunque disminuya la sensibilidad de la prueba. Sin embargo, en el caso de introducción de animales en áreas libres o protegidas con barreras sanitarias, el riesgo aumenta con cada nuevo animal que se quiera introducir. A medida que el tamaño del grupo aumenta, se incrementa también el riesgo de introducir uno o más individuos enfermos. Esta probabilidad se puede calcular mediante la siguiente fórmula (Marchevsky et al., 1989):

$$\text{Pr de incluir uno o más (E+/T-) en un grupo} = 1 - (\text{VP-})^n$$

Supongamos que queremos exportar un grupo de 100 bovinos provenientes del país recién mencionado a uno libre de la enfermedad y que para ello exigimos den negativos a la prueba tuberculínica. En este caso, y de acuerdo con lo visto en el punto «Valor predictivo negativo», tendremos que:

$$\begin{aligned} \text{VP-} &= \frac{(1 - 0,01) \times 0,89}{(1 - 0,01) \times 0,89 + 0,01 \times (1 - 0,91)} = \\ &= \frac{0,8811}{0,882} = 0,999 \text{ o } 99,9 \% \end{aligned}$$

Por lo que la posibilidad de que un animal esté sano habiendo dado negativo a la prueba (VP-) es cercana al 100 % (99,9 %).

Sin embargo, como esta probabilidad es individual para cada animal que da negativo a la prueba diagnóstica, y puesto que la probabilidad de que un animal esté sano

dado que dio negativo a la prueba diagnóstica es independiente del resultado de cualquier otro animal, debemos introducir el tamaño del grupo a importar. De esta forma vemos que:

$$\begin{aligned} \text{Pr de incluir uno o más (E+/T-)} &= 1 - (0,999)^{100} = \\ &= 0,095. \end{aligned}$$

Esto nos indica que si introducimos 100 animales negativos a la prueba de tuberculosis existe una probabilidad de casi el 10 % (9,5 %) de que al menos uno de ellos se encuentre enfermo de tuberculosis. En el Cuadro 14 se presentan varias combinaciones de sensibilidad de la prueba, prevalencia real de la enfermedad y tamaño del grupo. A modo de ejemplo, observemos que aun con una prueba de alta sensibilidad (99 %) y prevalencia relativamente baja (2 %) tenemos una probabilidad del 18,5 % de introducir uno o más individuo(s) enfermo(s) pero negativo(s) a la prueba si el grupo es de 1000 animales (Cuadro 14).

Cuadro 14. Probabilidad de incluir en un grupo uno o más individuos falsos negativos en varias combinaciones de sensibilidad, prevalencia real y tamaño del grupo

Sensibilidad (%)	Prevalencia real (%)	Tamaño del grupo			
		500	1000	2000	5000
90	2,0	63,9	87,0	98,3	100,0
	1,0	39,6	63,6	86,7	99,4
	0,5	22,2	39,5	63,4	91,9
95	2,0	39,9	63,9	87,0	99,4
	1,0	22,3	39,6	63,6	92,0
	0,5	11,8	22,2	39,5	71,5
99	2,0	9,7	18,5	33,5	64,0
	1,0	4,9	9,6	18,3	39,7
	0,5	2,5	4,9	9,6	22,2

Marchevsky (1989)

11. Efecto del tamaño de la muestra sobre la proporción de falsos negativos. Un ejemplo

Volviendo al ejemplo de la terapia selectiva de vaca seca (Tarabla, 1996) desarrollado en el punto «Sobre los falsos negativos» (Tarabla, 1995), en el Cuadro 15 se presenta el efecto del tamaño del rodeo sobre la probabilidad de incluir vacas con IIM dentro del grupo de animales negativos a los AMC o CMT (Tarabla, 1996).

Cuadro 15. Probabilidad (%) de incluir en un grupo una o más vacas falso negativas a los AMC o prueba de CMT en varias combinaciones de prevalencia real de infecciones intramamarias y tamaño del rodeo

Prueba	PR (%)	VP- (%)	Tamaño del rodeo		
			50	100	200
AMC	3,0	97,2	75,5	94,0	99,6
	5,0	95,4	90,7	99,1	99,9
	7,0	93,5	96,5	99,9	99,9
CMT	3,0	98,6	50,8	75,7	94,1
	5,0	97,6	69,9	90,9	99,8
	7,0	96,6	81,9	96,7	99,9

A modo de ejemplo, observemos que, aplicando el CMT con una prevalencia real de IIM del 5 %, el VP- es del 97,63 %, o sea que casi el 98 % de los animales negativos a la prueba está realmente sano. Si se efectúa tratamiento selectivo de la vaca seca en un rodeo de 50 animales, la probabilidad de incluir en el grupo de animales que no se van a tratar uno o más animales falsos negativos es de casi el 70 % (69,86 %). Sin embargo, si el tamaño del rodeo aumenta a 200 animales, aun manteniendo la prevalencia real de IIM (5 %) y el mismo criterio de selección (CMT), la probabilidad de incluir uno o más FN aumenta al 99 % (99,17 %).

El tratamiento selectivo de vaca seca es generalmente sugerido en aquellos rodeos con baja prevalencia real de IIM esgrimiendo razones económicas y otras propias del riesgo del uso de antibióticos en forma masiva. El VP- de toda prueba diagnóstica se mantiene en niveles altos cuando la prevalencia de la enfermedad es baja. Pero, a medida que el tamaño del grupo aumenta, se incrementa también el riesgo de incluir entre los animales negativos uno o más individuos enfermos falsos negativos (Tarabla, 1998). Sin que esto implique cambiar la decisión del uso de la terapia selectiva en rodeos con baja prevalencia de IIM, conviene entonces conocer que el riesgo de incluir uno o más animales FN en el grupo no tratado aumenta a medida que aumenta el tamaño del rodeo. Esto puede implicar incluso un cambio en la selección del antibiótico a utilizar, dado que en rodeos con baja prevalencia real de IIM pueden llegar a ser recomendados los antibióticos de corta acción para el tratamiento selectivo, mientras que en las estrategias donde se traten todos los animales del rodeo se deben utilizar los quimioterápicos de liberación lenta (Osteras et al., 1991).

El cálculo del valor predictivo de las pruebas diagnósticas es un elemento clave que permite comprender que positivo no es necesariamente enfermo, negativo no es sinónimo de sano y que estas probabilidades dependen no sólo del poder discriminatorio de la prueba diagnóstica sino también de la prevalencia real de la enfermedad. El VP+

está más influenciado por la especificidad de la prueba, mientras que el VP- está más relacionado a la sensibilidad de la misma. A medida que disminuye la prevalencia real decrece también el VP+ y aumenta la cantidad de diagnósticos falsos positivos. Como contrapartida, al incrementarse la prevalencia real hasta valores muy altos, disminuye el VP- y aumenta la cantidad de falsos negativos. Esta última cifra también se relaciona positivamente con la cantidad de individuos incluidos en el grupo de interés. Como se verá en el Capítulo 10, se pueden incrementar los VP utilizando combinación de pruebas. La ejecución en paralelo tiende a mejorar el VP-, mientras que la forma consecutiva incrementa el VP+.

Bibliografía

- Bayes, T. (1763).** An essay toward solving a problem in the doctrine of chance. *Philos. Trans. Roy. Soc.* **53**: 370–418.
- Coggins, L., Norcross, N.L., Nusbaum, S.R. (1972).** Diagnosis of equine infectious anemia by immunodiffusion test. *Am. J. Vet. Res.* **33**: 11–18.
- Edwards, D.S., Johnston, A.M., Mead, G.C. (1997).** Meat inspection: An overview of present practices and future trends. *Vet. J.* **154**: 135–147.
- Fleiss, J.L. (1981).** *Statistical Methods for Rates and Proportions*, 2nd Ed., John Wiley & Sons Inc., New York, 321 pp.
- Francis, J., Seiler, R.J., Wilkie, I.W., O’Byrne, D., Lumsden, M.J., Frost, A.J. (1978).** The sensitivity and specificity of various tuberculin tests using bovine PPD and other tuberculins. *Vet. Rec.* **103**: 420–435.
- Greiner, P.F., Mayewski, R.J., Mushlin, A.I., Greenland, P.G. (1981).** Selection and interpretation of diagnostic tests and procedures. *Ann. Int. Med.* **94**: 553–600.
- Lloyd-Webb, E.C., Campbell, P.H., Witt, D.J. (1995).** The specificity of the single cervical intradermal tuberculin test in a population of Tasmanian fallow deer putatively free of bovine tuberculosis. *Prev. Vet. Med.* **21**: 347–353.
- Marchevsky, N., Held, J.R., García-Carrillo, C. (1989).** Probability of introducing diseases because of false negative test results. *Am. J. Epidemiol.* **130**: 611–614.
- McClish, D., Quade, D. (1985).** Improving estimates of prevalence by repeated testing. *Biometrics.* **41**: 81–89.
- Osteras, O., Sandvik, L., Aursjo, G.G., Gjøl, G.G., Jorstad, A. (1991).** Assessment of strategy in selective dry cow therapy for mastitis control. *J. Vet. Med. B* **38**: 513–522.
- Rindsig, R.B., Rodewald, R.G., Smith, A.R., Thomsen, N.K., Spahr, S.L. (1979).** Mastitis history, California Mastitis Test, and somatic cell counts for identifying cows for treatment in a selective dry cow therapy program. *J. Dairy Sci.* **62**: 1335–1339.
- Slenning, B.D., Gardner, I.A. (1997).** Economic evaluation of risks to producers who use milk residue testing programs. *J. Am. Vet. Med. Ass.* **211**: 419–427.
- Tarabla, H.D., Calvino, L.F., Vitulich, C.A. (1994).** Encuesta sobre el uso de antibióticos en mastitis bovina. *Actas VII Cong. Arg. Cien. Vet., Buenos Aires (Capital)*, p. 354.
- Tarabla, H.D. (1995).** Factores que influyen el valor predictivo de una prueba diagnóstica. *Resúmenes 11 Congr. Arg. y 11 Congr. Latinoam. Zoonosis, Buenos Aires, Argentina*, p. 210.
- Tarabla, H.D. (1995).** Pruebas diagnósticas. Uso e interpretación epidemiológica. *En: La Brucelosis en Sistemas de Producción de Leche, INTA*: 1–19.
- Tarabla, H. D. (1996).** Efecto del tamaño del rodeo sobre la proporción de falsos negativos en el tratamiento selectivo de la vaca seca en rodeos con baja prevalencia de infecciones intramamarias. *Memorias Congr. Nac. Calidad de Leche y Mastitis, Río Cuarto, Córdoba*, pp. 18–20.
- Tarabla, H. D. (1996).** Importancia de la sensibilidad de la prueba diagnóstica y de la prevalencia de patógenos intramamarios en el tratamiento selectivo de la vaca seca. *Memorias Congr. Nac. Calidad de Leche y Mastitis, Río Cuarto, Córdoba*, pp. 21–23.
- Tarabla, H. D. (1998).** Algunas limitantes del tratamiento selectivo de la vaca seca (Some limitations of selective dry cow therapy). *Memorias Congr. Panam. Control de Mastitis y Calidad de Leche, Mérida, Yucatán, México*, pp. 283, 284.
- Yamamoto, R. (1975).** Characteristics an evaluation of screening tests. *Mycoplasmosis Workshop Am. Assoc. Avian Pathol. 112th Annu. Mtg. Am. Vet. Med. Assoc., Anaheim, California. Univ. California Davis, EPM Dept.*, 5 p.

Capítulo 9

Eficiencia

1. Cálculo de la eficiencia (Ef) de la prueba diagnóstica

Se denomina eficiencia a la habilidad de una prueba diagnóstica de detectar el verdadero estado de salud del individuo en una población compuesta por individuos enfermos tanto como sanos. Se define como:

$$Ef = P(T+ | E+) \times P(E+) + P(T- | E-) \times P(E-)$$

La fórmula anterior puede simplificarse de la siguiente forma:

$$Ef = \text{Sens} \times PR + \text{Esp} \times (1 - PR)$$

La eficiencia puede ser expresada como porcentaje multiplicando la fórmula anterior por 100.

La eficiencia describe para un determinado punto de corte y una prevalencia en particular qué fracción de todos los resultados obtenidos está compuesta por observaciones correctas. Supongamos que contamos con una población (n) de 10 000 animales, una prevalencia real del 10 %, y empleamos una prueba diagnóstica que posee una sensibilidad del 90 % y una especificidad equivalente al 95 %. En este caso la eficiencia será:

$$Ef = \{0,10 \times 0,90 + [(1 - 0,10) \times 0,95]\} \times 100$$
$$Ef = (0,09 + 0,855) \times 100 = \mathbf{94,5 \%}$$

Tomando como ejemplo la aplicación de una prueba inmunoenzimática (ELISA) comercial para el diagnóstico del Síndrome de Inmunodeficiencia Felina (SIF) con una sensibilidad del 95,7 % y una especificidad del 94,0%, (Galen y Gambino, 1975:237) si se aplica esta prueba en una población con una prevalencia real de SIF del 1% tendremos:

$$Ef = \{0,01 \times 0,957 + [(1 - 0,01) \times 0,94]\} \times 100$$

$$Ef = (0,00957 + 0,9306) \times 100 = \mathbf{94,0 \%}$$

Esto significa que un individuo que haya sido sometido a esta prueba para el diagnóstico de SIF en una población con una prevalencia real de la enfermedad del 1 % tiene un 94 % de chances de que el resultado obtenido refleje su verdadero estado de salud.

Utilizando el método de cálculo de los cuadros, tal como se hizo para los valores predictivos tendremos:

Cuadro 1. Eficiencia de la prueba diagnóstica

Condición	Enfermos	Sanos	Total
Test positivos	PV	FP	PV + FP
Test negativos	FN	NV	NV + FN
Total	PV + FN	NV + FP	PV + FP + NV + FN = n

Siendo la eficiencia equivalente a:

$$Ef = \frac{PV + NV}{n} \times 100$$

En el ejemplo anterior tendremos (Cuadro 2):

Cuadro 2. Cuadro de cálculo de la eficiencia de la prueba de ELISA para el diagnóstico de SIF en una población de 100 animales con una prevalencia real de la enfermedad del 1 %

Condición	Con SIF	Sin SIF	Total
ELISA positivos	957	5940	6897
ELISA negativos	43	93 060	93 103
Total	1000	99 000	100 000

Por lo tanto:

$$Ef = \frac{957 + 93\ 060}{100\ 000} \times 100 = \mathbf{94,0 \%}$$

La eficiencia tiene tres importantes limitaciones: (Altman et al., 1994):

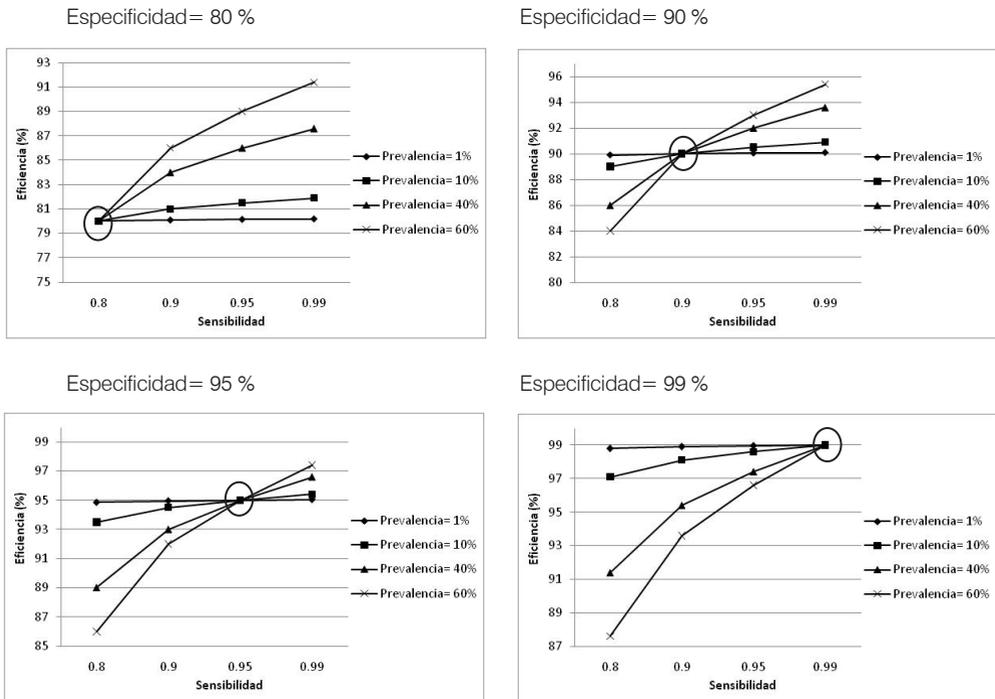
a) Es calculada generalmente para uno (o unos pocos) punto(s) de corte(s), por lo que puede llevar a conclusiones equivocadas.

- b) Es también altamente dependiente de la prevalencia de enfermedad y puede parecer alta, aun cuando la sensibilidad y la especificidad de la prueba sean pobres.
- c) Es definida como el porcentaje de resultados correctos, por lo que asume que el costo de los falsos positivos y los falsos negativos es el mismo. Con frecuencia esto no es verdad. Cuanto mayor sea la diferencia entre ambos costos, mayor será la distorsión que inducirá el cálculo de la eficiencia.

2. Efecto de la sensibilidad y especificidad de la prueba diagnóstica y la prevalencia real de la enfermedad sobre la eficiencia de la prueba

La eficiencia de una prueba diagnóstica es una función del poder discriminatorio de la misma y de la prevalencia real de la enfermedad. Salvo en casos de prevalencias muy altas (>50 %), las variaciones en la especificidad de la prueba tienen un impacto mayor en la eficiencia de la misma que las variaciones en la sensibilidad. En caso de superar el 50 % de prevalencia real, esta situación se invierte. Por otra parte, cuando la sensibilidad es idéntica a la especificidad, la eficiencia de la prueba se mantiene inalterable independientemente de la prevalencia de la enfermedad (Figura 1).

Figura 1. Eficiencia de una prueba diagnóstica en función de la sensibilidad, especificidad y prevalencia real



Nota: en el círculo se observa la eficiencia cuando la sensibilidad = especificidad.

En la Figura 1 se observa claramente que, mientras la sensibilidad es menor que la especificidad de la prueba diagnóstica, la eficiencia de la prueba es menor en aquellos casos donde la prevalencia real es alta. Cuando la sensibilidad = especificidad, la eficiencia es igual independientemente de la prevalencia de la enfermedad. Por último, a medida que el valor de sensibilidad se torna mayor que el de especificidad, la prueba se hace más eficiente en cuanto aumenta la prevalencia real de la enfermedad.

En un contexto en el cual la prevalencia de la enfermedad disminuye (ej.: en el caso de un programa de erradicación) la eficiencia de la prueba también decrece si su sensibilidad es mayor que su especificidad. Si por el contrario la prevalencia de la enfermedad aumenta (ej.: por la introducción de animales enfermos), la eficiencia de la prueba disminuye en el caso en que la especificidad sea mayor que la sensibilidad (Trajstman, 1979). En términos generales, en medicina veterinaria el concepto de eficiencia no tiene relevancia dado que rara vez adquieren idéntica importancia un resultado falso positivo y un resultado falso negativo. En medicina humana, sin embargo, puede ser un elemento a considerar en casos de enfermedades serias pero tratables en las que si un caso no se diagnostica puede causar daño irreparable, pero igual daño puede causar que se efectúe el diagnóstico positivo con el consecuente tratamiento cuando la enfermedad esté ausente (ej.: infarto de miocardio, diabetes) (Galen y Gambino, 1975:237). Sin embargo, se debe tener en cuenta que tanto cuando la prevalencia real es baja como cuando es alta es posible tener una alta eficiencia simplemente como resultado de haber clasificado correctamente a uno de los dos grupos enfermos o sanos. Más se aparta la prevalencia de enfermedad del 50%, más influencia tiene esto en la eficiencia de las pruebas.

A modo de ejemplo, supongamos una prueba altamente específica (99,0 %) y, a su vez, con idéntica sensibilidad a la que habríamos obtenido por mera acción del azar (50 %) aplicada a una población con una prevalencia real de enfermedad del 0,1 % (Cuadro 3).

Cuadro 3. Eficiencia de una prueba diagnóstica (Sens 50,0 %, Esp 99,0 %) aplicada en una población de 100 000 animales con una prevalencia real de enfermedad del 0,1 %

Condición	Enfermos	Sanos	Total
Prueba positiva	50	999	1049
Prueba negativa	50	98 901	98 951
Total	100	99 900	100 000

$$Ef. = \frac{50 + 98\ 901}{100\ 000} \times 100 = 99,0 \%$$

Es decir que esta prueba diagnóstica aplicada en esas circunstancias tendría una probabilidad de diagnósticos correctos del 99,0 %, aunque su habilidad para detectar animales verdaderamente positivos entre los enfermos no supera lo que habríamos obtenido «tirando la moneda».

Cuando se utilizan las pruebas diagnósticas para estimar la prevalencia de las enfermedades, esta estimación es menos exacta a medida que disminuye la sensibilidad y especificidad de la prueba (ej.: disminuye su eficiencia). Por otra parte, cuando la prevalencia es calculada como un intervalo, más que como un valor único, a medida que aumenta la eficiencia de la prueba disminuye el intervalo de la estimación (Marchevsky, 1974).

Bibliografía

- Altman, D.G., Bland, J.M. (1994).** Diagnostic tests, 2: Predictive values. *B. Med. J.* **309**: 102.
- Galen, R. S. and Gambino, S. R. (1975).** Beyond Normality, the Predictive Value and Efficiency of Medical Diagnoses. John Wiley & Sons, New York, 237 pp.
- Gardner, I.A., Greiner, M. (1999).** Advanced Methods for test validation and interpretation in Veterinary Medicine. Freie Univ. Berlin, 74 pp.
- Marchevsky, N. (1974).** Errores en las estimaciones de prevalencia en estudios de población: Un método para calcular la prevalencia real. *Zoonosis.* **16**: 81–109.
- Smith, R.D. (2005).** *Veterinary Clinical Epidemiology* – 3^o Edición. Taylor & Francis, Boca Raton, Florida (Estados Unidos de América). 259 pp.
- Trajstman, A.C. (1979).** Diagnostic tests, sensitivity, specificity, efficiency and prevalence. *Aust. Vet. J.* **55**: 501.

Capítulo 10

Combinación de pruebas diagnósticas

Una práctica muy frecuente es la utilización de varias pruebas para emitir un diagnóstico. En estos casos, la interpretación de los resultados depende de la secuencia en la cual las pruebas son ejecutadas y de la forma en que se integran sus respectivos resultados (Smith, 1991:225). En el caso de utilizar dos pruebas existen dos posibilidades básicas, la primera es efectuar las pruebas en forma paralela y la segunda es realizarlas en forma secuencial, procesando en la siguiente sólo aquellas que dieron positivo a la prueba anterior. Obviamente, las posibilidades de diseño de pruebas múltiples no se agotan en estos dos ejemplos.

Estas estrategias parten del supuesto de que la segunda prueba arroja información adicional superior a la proporcionada por la anterior dado que, de no ser así, el rendimiento de ambas modalidades no será el esperado (Tarabla, 2000). Al diseñar una estrategia de combinación de pruebas diagnósticas no sólo se debe conocer la sensibilidad y especificidad de las mismas sino también si los fenómenos que ellas miden son diferentes o independientes y si hay formas o etapas de la enfermedad que no son detectadas por las pruebas (Riegelman y Hirsch, 1989:256). Aquellas pruebas que miden similares procesos biológicos, como la respuesta inmune (anticuerpos) sérica frente a agentes infecciosos, deberían generar resultados similares y condicionados por el verdadero estatus sanitario del animal.

Si no se cumple el supuesto de independencia entre las pruebas, los cálculos de la probabilidad de enfermedad a partir de los resultados de varias pruebas tenderán a sobreestimar el valor real de la estrategia (Smith, 1991:225). Una tercera alternativa, utilizada en los planes de control y erradicación que impliquen la eliminación de los reaccionantes positivos, es la repetición de la misma prueba a intervalos regulares en los animales reaccionantes negativos en la oportunidad anterior.

Supongamos que en una población de 100 000 animales con una enfermedad que tiene una prevalencia real del 5 % efectuamos dos pruebas diagnósticas (A y B) cuyos resultados hipotéticos se presentan en el Cuadro 1.

Cuadro 1. Datos hipotéticos luego de la ejecución de dos pruebas diagnósticas en una población de 100 000 animales con una prevalencia real de enfermedad del 5 %

Condición	Enfermos	Sanos	Total
A+ B+	3800	1425	5225
A+ B-	950	12 825	13 775
A- B+	200	8075	8275
A- B-	50	72 675	72 725
Total	5000	95 000	100 000

La sensibilidad de la prueba A es del 95 %

$$\frac{3800 + 950}{5000} \times 100,$$

la sensibilidad de la prueba B es del 80 %

$$\frac{3800 + 200}{5000} \times 100,$$

la especificidad de la prueba A es del 85 %

$$\frac{8075 + 72 675}{95 000} \times 100$$

la especificidad de la prueba B es del 90 %

$$\frac{12 825 + 72 675}{95 000} \times 100,$$

Por otra parte, los valores predictivos bajo esta prevalencia real de enfermedad del 5 % se detallan a continuación:

El valor predictivo positivo utilizando la prueba A es del 25 %

$$\frac{3800 + 950}{5225 + 13 775} \times 100,$$

el valor predictivo positivo utilizando la prueba B es del 29,6%

$$\frac{3800 + 200}{5225 + 8275} \times 100,$$

el valor predictivo negativo con la prueba A es del 99,7 %

$$\frac{8075 + 72\ 675}{8275 + 72\ 725} \times 100,$$

mientras que el valor predictivo negativo con la prueba B es del 98,8 %

$$\frac{12\ 825 + 72\ 675}{13\ 775 + 72\ 725} \times 100$$

1. Ejecución de pruebas en paralelo

Supongamos que aplicamos las pruebas A y B en forma paralela y decidimos clasificar como enfermos a los reaccionantes positivos a cualquiera de las dos pruebas (Galen y Gambino, 1975:237). Los resultados posibles serán los siguientes:

A+ B+ = +

A+ B- = +

A- B+ = +

A- B- = -

De acuerdo con los resultados presentados en el Cuadro 1, si efectuamos el diagnóstico de esta manera daremos como positivos dentro de los 5000 animales verdaderamente enfermos:

A+ B+ = 3800

A+ B- = 950

A- B+ = 200

Esto puede ser resumido en el Cuadro 2.

Cuadro 2. Pruebas paralelas

Condición	Enfermos	Sanos	Total
A o B positivos	4950	22 325	27 275
A y B negativos	50	72 675	72 725
Total	5000	95 000	100 000

Por lo tanto, la sensibilidad de esta estrategia es del 99,0 %

$$\frac{(3800 + 950 + 200)}{5000} \times 100,$$

Por otra parte, dentro de los 95 000 animales verdaderamente sanos hay 72 675 clasificados correctamente por la combinación A- B-, por lo que la especificidad es del 76,5 %

$$\frac{72\ 675}{95\ 000} \times 100$$

Nótese que, en este caso, con una prevalencia real del 5 %, se obtiene un alto VP-

$$\frac{72\ 675}{72\ 725} \times 100 = 99,9 \%$$

pero el VP+ es de sólo el 18,1 %

$$\frac{4950}{27\ 275} \times 100,$$

Con la ejecución de las pruebas paralelas incrementamos la sensibilidad de nuestra prueba por encima de las sensibilidades de cada una de ellas en detrimento de la especificidad del diagnóstico. En el ejemplo presentado (Cuadros 1 y 2), las sensibilidades de ambas pruebas son condicionalmente independientes (Galen y Gambino, 1975:237), es decir, la sensibilidad de la prueba B no está influenciada por los resultados de la primera (A) en la población de individuos enfermos (Cuadro 3).

Cuadro 3. Pruebas A y B en 5000 animales enfermos asumiendo independencia condicional entre sus sensibilidades (Sens A= 95 %, Sens B= 80 %)

Prueba	B+	B-	Total
A+	3800	950	4750
A-	200	50	250
Total	4000	1000	5000

En este caso, la sensibilidad de la prueba B entre los positivos a la primera prueba (A) es del 80 % tanto entre los positivos a la misma

$$\frac{3800}{4750} \times 100,$$

como entre los negativos

$$\frac{200}{250} \times 100$$

No obstante, si ambas pruebas están biológicamente asociadas (ej.: si miden el mismo tipo de respuesta en el mismo estadio de la enfermedad) la ganancia en la sensibilidad será mínima (Riegelman y Hirsch, 1989:256). En el Cuadro 4 se presentan los mismos datos hipotéticos pero asumiendo en este caso, dependencia absoluta entre las sensibilidades de ambas pruebas.

Cuadro 4. Pruebas A y B en 5.000 animales enfermos asumiendo dependencia condicional absoluta entre sus sensibilidades (Sens A= 95 %, Sens. B= 80 %)

Prueba	B+	B-	Total
A+	4000	750	4750
A-	0	250	250
Total	4000	1000	5000

Asumiendo dependencia condicional absoluta entre las sensibilidades de ambas pruebas, la segunda (B) no detecta nuevos positivos entre los que dieron negativos a la primera (A). La sensibilidad de B es ahora del 84,2 % en los positivos a la prueba A

$$\frac{4000}{4750} \times 100,$$

y del 0 % entre los negativos

$$\frac{0}{250} \times 100$$

Por ello, la sensibilidad de la estrategia permanece en un 95,0 %, dado que:

$$\begin{aligned} A+ B+ &= 4000 \\ A+ B- &= 750 \\ A- B+ &= 0 \end{aligned}$$

$$\frac{4000 + 750 + 0}{5000} \times 100 = 95 \%$$

En este ejemplo de dependencia absoluta entre las sensibilidades, el VP- de la estrategia en paralelo no se verá incrementado con respecto al que habríamos obtenido utilizando sólo la prueba A. De esto se desprende que la ganancia en sensibilidad y VP- estará en íntima relación con el grado de dependencia entre las sensibilidades de ambas pruebas.

La estrategia en paralelo tiene como efecto final demostrar que el individuo no está enfermo (Smith, 1991:225) y funciona bien cuando cada prueba detecta una etapa distinta de la enfermedad, aunque las sensibilidades de las pruebas individuales no sean especialmente altas (Riegelman y Hirsch, 1989:256). A veces una prueba puede detectar un estadio temprano y otra más avanzado. En otros casos, una prueba puede detectar una enfermedad de evolución rápida o agresiva mientras que la segunda detecta una de progresión gradual o lenta. Cuando se requiera minimizar los diagnósticos falsos negativos se puede cambiar la prueba en uso por otra más sensible, cambiar el punto de corte de la prueba en uso o utilizar esta estrategia.

2. Ejecución de pruebas consecutivas

Supongamos que aplicamos las dos pruebas en forma consecutiva (también llamada en serie) (Galen y Gambino, 1975:237) y decidimos clasificar como enfermos a los individuos que, habiendo sido positivos a una prueba, son también positivos a la segunda. Los resultados posibles son los siguientes:

- La prueba A es aplicada primero y todos los positivos son re-analizados con la prueba B (A+ B+ = +).
- La prueba B es aplicada primero y todos los positivos son nuevamente analizados con la prueba A (B+ A+ = +).

En ambos casos, de los 5000 animales enfermos 3800 estarán correctamente clasificados como tales, por lo que la sensibilidad de la combinación será del 76,0 %. Por otra parte, todas las otras combinaciones (A+ B-, A- B+ y A- B-) serán dadas como negativos. De acuerdo con el ejemplo del Cuadro 1 los animales sanos así clasificados son respectivamente 8075, 12 825 y 72 675. Por consiguiente, la especificidad será del 98,5 %:

$$\frac{8075 + 12\ 825 + 72\ 675}{95\ 000} \times 100$$

Con la ejecución de las pruebas en forma consecutiva incrementamos la especificidad de la prueba por encima de las especificidades de cada una de ellas pero se disminuye la sensibilidad del diagnóstico.

A continuación se verá qué ocurre con el valor predictivo positivo si alteramos el orden de ejecución de las pruebas. Conforme a los resultados presentados en el Cuadro 1, si efectuamos la prueba A primero tendremos (Cuadro 5):

Cuadro 5. Prueba A (Sens = 95 %, Esp = 85 %)

Condición	Enfermos	Sanos	Total
Test positivos	4750	14 250	19 000
Test negativos	250	80 750	81 000
Total	5000	95 000	100 000

Si posteriormente aplicamos la prueba B a todos los animales positivos a la prueba A, tendremos (Cuadro 6):

Cuadro 6. Prueba B (Sens = 80 %, Esp = 90 %) luego de la Prueba A

Condición	Enfermos	Sanos	Total
Test positivos	3800	1425	5225
Test negativos	950	12 825	13 775
Total	4750	14 250	19 000

Nótese que la aplicación de la segunda prueba redujo el número de falsos positivos de 14 250 a 1425, y aumentó por consiguiente el VP+. En este caso, el VP+ es del 72,7 %, dado que 3800 de los 5225 animales positivos a esta segunda prueba están verdaderamente enfermos. Esto mejora notablemente el VP+ que se obtuvo empleando las pruebas en forma paralela (18,1 %) o el que se habría obtenido utilizando la prueba A solamente, que era del 25 %,

$$\frac{4750}{19\ 000} \times 100 = 25\ %$$

Sin embargo, 250 individuos enfermos fueron declarados negativos luego de la primera prueba, los que, sumados a los 950 obtenidos después de efectuada la prueba B, suman 1200 falsos negativos, por lo que el valor predictivo negativo experimenta un leve descenso:

$$\frac{80\ 750 + 12\ 825}{81\ 000 + 13\ 775} \times 100 = 98,7\ %$$

Si optamos, en cambio, por efectuar la prueba B, primero tendremos (Cuadro 7):

Cuadro 7. Prueba B (Sens = 80 %, Esp = 90 %)

Condición	Enfermos	Sanos	Total
Test positivos	4000	9500	13 500
Test negativos	1000	85 500	86 500
Total	5000	95 000	100 000

Si luego reanalizamos con la prueba A los animales que dieron positivo a la prueba B tendremos (Cuadro 8):

Cuadro 8. Prueba A (Sens = 95 %, Esp = 85 %) luego de la prueba B.

Condición	Enfermos	Sanos	Total
Test positivos	3800	1425	5225
Test negativos	200	8075	8275
Total	4000	9500	13500

Nótese que el VP+ es también del 72,7 %:

$$\frac{3800}{5225} \times 100 = 72,7 \%$$

que implica una sensible mejora en el VP+ del 29,6 % obtenido al utilizar la prueba B solamente,

$$\frac{4000}{13 500} \times 100 = 21,9 \%$$

Por otra parte, el VP- es nuevamente del 98,7 %:

$$\frac{85 500 + 8075}{86 500 + 8275} \times 100 = 98,7 \%$$

La estrategia consecutiva tiene como efecto final demostrar que el individuo está verdaderamente enfermo de la afección de interés (Smith et al., 1986). Un dato interesante es que, *a priori*, no importa la secuencia en la ejecución de ambas pruebas ya que los resultados finales de la estrategia son los mismos. Ahora bien, si juzgamos la elección de la secuencia de estas pruebas por el solo hecho del número de muestras a ejecutar, convendría efectuar primero la prueba B, dado que deberíamos hacer 113

500 determinaciones (100 000 + 13 500) *versus* 119 000 en el caso opuesto (100 000 + 19 000), con exactamente la misma cantidad de verdaderos positivos detectados (3800). El VP+ y el VP- no están influenciados por la elección de la primera prueba, siempre asumiendo que ambas pruebas son independientes y, por lo tanto, que la que se realiza en segundo lugar no cambia su sensibilidad y especificidad, cualquiera sea el resultado obtenido en la primera. Sin embargo, para la elección de la secuencia de las pruebas se tienen en cuenta además otros factores de importancia ya mencionados en el punto «Características deseables en una prueba diagnóstica (economía, disponibilidad de materiales, simplicidad, etc.)». Si todos estos factores son idénticos en ambas pruebas, se debe utilizar primero la prueba más específica para reducir el número de animales falsos positivos y, por ende, disminuir el número de muestras que deben ser procesadas por la segunda (Smith, 1991:225). No obstante, habitualmente son razones de seguridad y de costo las que determinan qué prueba se realizará primero (Riegleman y Hirsch, 1989:256). La tendencia es utilizar en principio pruebas de buena sensibilidad que resulten baratas, rápidas y sencillas (prueba tamiz) para luego procesar las muestras que resulten positivas por otra prueba altamente específica (confirmatoria) aunque más cara o complicada. Por otra parte, en los estadios iniciales de un plan de control generalmente se utiliza una sola prueba diagnóstica, ya que cuando existe una alta prevalencia de la enfermedad el VP+ se mantiene relativamente alto. En estas condiciones, una prueba tamiz con alta sensibilidad y aceptable especificidad permitirá detectar rápidamente los animales enfermos positivos para tomar con ellos medidas que eviten la continua propagación de la enfermedad. Es en las etapas finales de un plan de erradicación cuando más importancia adquiere la estrategia consecutiva puesto que aumenta el VP+ y minimiza los reaccionantes falsos positivos. En estos casos se puede cambiar la prueba en uso por otra más específica, cambiar el punto de corte de la prueba en uso o utilizar esta estrategia.

Es factible que las pruebas consecutivas lleven más tiempo que la ejecución en paralelo, pero el incremento de la especificidad del diagnóstico puede permitir descartar la enfermedad empleando menos pruebas. La estrategia consecutiva también puede ser de ayuda cuando ambas pruebas tienen baja especificidad (Riegelman y Hirsch, 1989:256).

En el caso anterior se asume independencia condicional entre las especificidades (Cuadro 9).

Cuadro 9. Pruebas A y B en 95 000 animales sanos asumiendo independencia condicional entre sus especificidades (Esp A= 85 %, Esp B= 90 %)

Prueba	B+	B-	Total
A+	1425	12 825	14 250
A-	8075	72 675	80 750
Total	14 250	80 750	95 000

Según la especificidad de B es del 90 % tanto entre los positivos a la prueba A

$$\frac{12\ 825}{14\ 250} \times 100$$

como entre los negativos

$$\frac{72\ 675}{80\ 750} \times 100$$

No obstante, de existir algún grado de dependencia condicional entre ambas especificidades, la segunda prueba detectará una menor cantidad de individuos negativos entre los positivos a la primera (Francis et al., 1978). El razonamiento es similar al presentado con respecto a las sensibilidades. Consecuentemente, el incremento en especificidad y VP+ de una estrategia consecutiva estará relacionado con el grado de dependencia relativa entre las especificidades de las pruebas utilizadas. Cuanto mayor sea la dependencia, menor será la ganancia.

3. Repetición de la misma prueba

Esta modificación de la interpretación de las pruebas consecutivas consiste en aplicar la misma prueba a los animales que arrojaron resultados negativos en la primera oportunidad. El efecto final es similar a la interpretación paralela, es decir, probar que el individuo no está enfermo de la afección de interés. Los resultados posibles con esta estrategia serán los siguientes:

A+ = +

A- = Repetir prueba A.

o bien,

B+ = +

B- = Repetir prueba B.

Supongamos que aplicamos la prueba A clasificando como enfermos a los reaccionantes positivos (Cuadro 10).

Cuadro 10. Prueba A (Sens = 95 %, Esp = 85 %)

Condición	Enfermos	Sanos	Total
Test positivos	4750	14 250	19 000
Test negativos	250	80 750	81 000
Total	5000	95 000	100 000

Si posteriormente repetimos la prueba A a todos los animales negativos en el primer diagnóstico (Cuadro 11).

Cuadro 11. Prueba A (Sens = 95 %, Esp = 85 %) luego de la Prueba A

Condición	Enfermos	Sanos	Total
Test positivos	238	12 113	12 351
Test negativos	12	68 637	68 647
Total	250	80 750	81 000

Por lo tanto, la sensibilidad, especificidad y valores predictivos de esta estrategia son:

Sensibilidad del 99,8 %

$$\frac{4750 + 238}{5000} \times 100,$$

la especificidad del 72,2 %

$$\frac{68 637}{95 000} \times 100,$$

el valor predictivo positivo del 15,9 %

$$\frac{4750 + 238}{19 000 + 12 351} \times 100,$$

y el valor predictivo negativo del 99,98 %

$$\frac{68 637}{68 649} \times 100$$

Si hubiéramos utilizado la prueba B en esta estrategia, estos valores serían del 96,0, 81,1, 21 y 99,7 % para la sensibilidad, especificidad, VP+ y VP-, respectivamente. Esta repetición de la prueba en el mismo individuo es la base de los programas de control basados en la aplicación de pruebas diagnósticas y eliminación de los reaccionantes positivos (ej.: brucelosis y tuberculosis). En este caso se debe tener en cuenta que si los animales positivos son continuamente enviados a descarte, el rodeo será eventualmente devastado (Riegelman y Hirsch, 1989:256). Sin embargo, al haber dependencia

condicional debido a que es la misma prueba que se repite, un animal que dio falso negativo en un primer análisis es probable que nuevamente repita este resultado al ser analizado por segunda vez (Gardner y Greiner, 1999:74). Es por ello que los planes de control de brucelosis y tuberculosis en bovinos por ejemplo, contemplan la realización de dos pruebas consecutivas (BPA y fijación del complemento, intradermorreacción en pliegue anocaudal y comparativa con PPD aviar en tabla del cuello) para incrementar la especificidad del diagnóstico, aumentar el valor predictivo de la prueba positiva y disminuir la cantidad de falsos positivos.

4. Comparación entre las estrategias paralela, consecutiva y de repetición

Si comparamos la sensibilidad, especificidad y los valores predictivos de cada estrategia hasta aquí desarrollada, las consecuencias de la aplicación de cada uno de estos criterios serán (Cuadro 12):

Cuadro 12. Utilización de la prueba A (Sens = 95 %, Esp = 85 %) y la prueba B (Sens = 80 %, Esp = 90 %) con distintas estrategias y una prevalencia real del 5 %

Estrategia	Sensibilidad	Especificidad	VP+	VP-
Prueba A	95,0	85,0	25,0	99,7
Prueba B	80,0	90,0	29,6	98,8
Paralela	99,0	76,5	18,1	99,9
Consecutiva	76,0	98,5	72,7	98,7
Repetición A	99,8	72,2	15,9	99,9
Repetición B	96,0	81,0	21,0	99,7

Como resumen, en el Cuadro 13 se exponen de manera comparada los efectos de emplear las diferentes estrategias de combinación de pruebas diagnósticas.

Cuadro 13. Resultados de la aplicación de las estrategias paralela, consecutiva y de repetición

Resultado	Estrategia		
	Paralela	Consecutiva	Repetición
Sensibilidad	↑	↓	↑
Especificidad	↓	↑	↓
Valores predictivos	↑ VP-	↑ VP+	↑ VP-
Objetivo	Descartar la presencia de una enfermedad.	Confirmar la presencia de una enfermedad.	Descartar la presencia de una enfermedad.
Resultado	↓ Falsos negativos	↓ Falsos positivos	↓ Falsos negativos
Tiempo de ejecución	Menor	Mayor	Mayor
Uso más frecuente	Tamizado, emergencias.	Control y erradicación, confirmación de un diagnóstico.	Control y erradicación.

Independientemente de la estrategia que se desee emplear (pruebas en serie y en paralelo), se supone que la segunda prueba proporciona una información adicional superior a la que proporciona la primera. Obviamente, cuanto menor sea la dependencia condicional existente entre las pruebas utilizadas mejores serán los resultados de cada estrategia.

5. Estrategias paralela, consecutiva y de repetición. Ejemplos

5.1. Estrategia paralela

La utilización del criterio paralelo es más frecuente en medicina humana que en veterinaria. En un tamizado para detección precoz de enfermedades serias, las pruebas se utilizan en paralelo (ej.: palpación y mamografía en cáncer de mama) dado que es muy importante tener una alta sensibilidad para poder detectar a todo individuo enfermo. En el seguimiento intensivo de los casos individuales positivos se utilizan posteriormente pruebas complementarias (ej.: biopsia) en una estrategia consecutiva al tamizado para descartar los diagnósticos falsos positivos.

En medicina veterinaria, el diagnóstico de tuberculosis bovina se basa en la intradermorreacción (IR) con tuberculina. Sin embargo, los inconvenientes operativos que conlleva la lectura de los resultados a las 72 horas hacen que continúen los esfuerzos de los investigadores por encontrar una prueba serológica que la reemplace. La

inmunidad celular detectada por la reacción intradérmica predomina en las etapas tempranas de la enfermedad, mientras que los anticuerpos séricos no son detectados hasta etapas más tardías de la evolución de la infección (Lightbody et al., 1998). La combinación en paralelo de la IR y la prueba inmunoenzimática anamnésica (ELISA A) ha sido recomendada para la erradicación de tuberculosis caprina (Gutiérrez et al., 1998). En esta especie, la IR tiene una sensibilidad del 80 % y una especificidad del 99,1 % (Sanson, 1988). En el caso del ELISA A, estas cifras alcanzan el 88,6 % y el 95,8 %, respectivamente (Gutiérrez et al., 1998).

A modo de ejemplo, supongamos que efectuamos ambas pruebas en una población de 100 000 caprinos que tiene una prevalencia real de tuberculosis del 10 %. Los hipotéticos resultados, asumiendo independencia entre ambas pruebas, se presentan en el Cuadro 14.

Cuadro 14. Ejecución de la intradermorreacción (IR, Sens = 80,0 %, Esp = 99,1 %) y la prueba inmunoenzimática anamnésica (ELISA A, Sens = 88,6 %, Esp = 95,8 %) en una población de 100 000 animales con una prevalencia real de enfermedad del 10 %

Condición	Con tuberculosis	Sin tuberculosis	Total
IR+ ELISA A+	709	3	712
IR+ ELISA A-	91	78	169
IR- ELISA A+	177	375	552
IR- ELISA A-	23	8544	8567
Total	1000	9000	10 000

De acuerdo con los resultados presentados en el Cuadro 14, si los interpretamos en paralelo daremos como positivos dentro de los 1000 animales tuberculosos a:

IR+ ELISA A+ = 709

IR+ ELISA A- = 91

IR- ELISA A+ = 177

Por lo tanto, la sensibilidad aumentó al 97,7 %

$$\frac{(709 + 91 + 177)}{1000} \times 100$$

Esta cifra es mayor a las sensibilidades de cada prueba ejecutadas en forma individual y asegura un alto VP- (99,7 %). En este ejemplo, las posibilidades de dependencia condicional entre pruebas (Gardner y Greiner, 1999:74) son relativamente bajas puesto que detectan distintas respuestas inmunológicas en etapas diferentes de la enfermedad. La habilidad del ELISA A para detectar cabras con lesiones severas que

resultan negativas a la IR puede ayudar a controlar rápidamente la tuberculosis (Gutiérrez et al., 1998). Obviamente, la contrapartida es el aumento de la probabilidad de falsos positivos al disminuir la especificidad del diagnóstico (94,9 %) y, consecuentemente, el VP+ (68,2 %).

5.2. Estrategia consecutiva.

Veamos ahora un ejemplo utilizando la prueba en placa con antígeno tamponado como prueba tamiz (Sens = 98,7 %, Esp = 95,4 %) que es el test de aglutinación para diagnóstico de brucelosis con la sensibilidad más alta. Como prueba confirmatoria utilizaremos la de fijación del complemento (Sens = 92,9 %, Esp = 99,7 %) (Dohoo et al., 1986).

Luego de ejecutar las pruebas del antígeno tamponado cuando la prevalencia real de brucelosis es del 10 %, tendremos (Cuadro 15):

Cuadro 15. Prueba del antígeno tamponado (BPA, Sens = 98,7 %, Esp = 95,4 %) en una población de 100 000 animales con una prevalencia real de brucelosis del 10 %

Condición	Con brucelosis	Sin brucelosis	Total
BPA positivos	9870	4140	14 010
BPA negativos	130	85 860	85 990
Total	10 000	90 000	100 000

Si posteriormente aplicamos la prueba de fijación del complemento a todos los animales positivos a la prueba tamiz, tendremos (Cuadro 16):

Cuadro 16. Prueba de fijación del complemento (FC, Sens = 92,9 %, Esp. = 99,7 %) en los animales positivos a la prueba del antígeno tamponado

Condición	Con brucelosis	Sin brucelosis	Total
FC positivos	9169	12	9181
FC negativos	701	4128	4829
Total	9870	4140	14 010

En este caso, el VP+ es de

$$\frac{9169}{9181} \times 100 = 99,9 \%$$

comparado con el 70,4 % que habríamos obtenido al utilizar la prueba de aglutinación solamente. Sin embargo, de existir algún grado de dependencia condicional entre las especificidades de ambas pruebas (que es el escenario más probable), la ganancia en especificidad y VP+ será menor a lo expresado (Gardner y Greiner, 1999:74).

5.3. Estrategia de repetición

Supongamos que aplicamos la prueba tuberculínica simple en el pliegue anocaudal con PPD (Francis et al., 1978) y clasificamos como tuberculosos a los reaccionantes positivos (Cuadro 17).

Cuadro 17. Intradermorreacción (IR) en pliegue anocaudal con PPD (Sens = 81,8 % y Esp = 96,3 %) en una población de 100 000 bovinos con una prevalencia real de tuberculosis del 5 %

Condición	Con tuberculosis	Sin tuberculosis	Total
IR positivos	4090	3515	7605
IR negativos	910	91 485	92 395
Total	5000	95 000	100 000

En un plan de control, los animales positivos habrían sido eliminados de los rodeos. Al repetir la misma prueba a todos los animales restantes se obtendrá (Cuadro 18):

Cuadro 18. Repetición de la intradermorreacción (IR) en pliegue anocaudal (Sens = 81,8 %, Esp = 96,3 %)

Condición	Con tuberculosis	Sin tuberculosis	Total
IR positivos	744	3385	4129
IR negativos	166	88 100	88 266
Total	910	91 485	92 395

Por lo tanto, con este criterio la sensibilidad es del 96,7 %, la especificidad del 92,7 %, el VP+ del 41,2 % y el VP- del 99,8 %. A medida que se avanza en el plan de control, menor será la probabilidad de que un animal positivo a la prueba esté tuberculoso. Cuando se llega a esas instancias se requiere el uso de pruebas complementarias ejecutadas en forma consecutiva (ej.: con PPD aviar) para aumentar la especificidad del diagnóstico.

Por otra parte, la repetición de una prueba sobre los mismos animales resulta en una dependencia condicional entre las sensibilidades de cada ejecución. Si un animal estaba infectado crónicamente y fue negativo a la prueba, los muestreos subsecuentes también tenderán a ser falsos negativos (Gardner y Greiner, 1999:74).

6. Dependencia condicional

Llegados a este punto, podemos pensar que virtualmente es factible alcanzar el 100 % de sensibilidad si aplicamos en paralelo dos o tres pruebas diagnósticas o el 100 % de especificidad con tres o cuatro pruebas diagnósticas usadas en serie. Este razonamiento resulta lógico siempre y cuando asumamos que, en individuos enfermos, la probabilidad de que una segunda prueba genere un resultado positivo es la misma en muestras que fueron positivas o negativas a una prueba anterior. Obviamente es preciso considerar un supuesto similar para los individuos sanos.

No obstante, en la realidad es prácticamente imposible alcanzar esos valores de sensibilidad y especificidad empleando pruebas de manera combinada. Esto se debe a que las pruebas utilizadas en combinación están con frecuencia correlacionadas entre sí. Cuanto mayor sea la correlación entre las pruebas, menor información adicional se obtiene al usar dichas pruebas en combinación. Por ejemplo, si dos pruebas diagnósticas están basadas en la identificación de la respuesta humoral ante un agente patógeno, si la prueba 1 dio un resultado negativo, es muy probable que la prueba 2 también genere un resultado negativo. En cambio, cuando se emplean pruebas que miden aspectos diferentes de una enfermedad o cuando se utilizan diferentes métodos diagnósticos (ej.: pruebas serológicas y de imagen) se incrementa el aporte de información útil para el diagnóstico.

Vamos a profundizar en lo anterior basándonos en un ejemplo hipotético. Supongamos que queremos determinar el estatus sanitario de una población de animales con respecto a una determinada enfermedad. Para ello aplicamos dos pruebas diagnósticas. La prueba A tiene una sensibilidad del 95 % y una especificidad del 85 %, mientras que la prueba B posee una sensibilidad y especificidad del 80 %.

Para analizar el efecto de la dependencia en la sensibilidad de las pruebas diagnósticas supongamos que aplicamos ambas pruebas en paralelo. De esta forma, el resultado se presenta en el cuadro 19.

Cuadro 19. Independencia condicional en la sensibilidad de dos pruebas diagnósticas aplicadas en paralelo

Condición	Enfermos	Sanos	Total
A+ B+	76	30	106
A+ B-	19	120	139
A- B+	4	170	174
A- B-	1	680	681
Total	100	1000	1100

Si queremos comprobar que la sensibilidad de las dos pruebas empleadas son independientes entre sí, debemos observar las siguientes condiciones:

- a) La sensibilidad de la segunda prueba no depende de que el resultado de la primera haya sido positivo o negativo en la población enferma.
- b) La segunda prueba detecta todos los animales enfermos que su sensibilidad le permite entre aquellos que resultaron negativos a la primera.

Realizamos ahora los cálculos y podemos observar que:

$$\text{Sensibilidad Prueba B en individuos A+} = 76 / (76+19) = 0,8 \text{ o } 80 \%$$

$$\text{Sensibilidad Prueba B en individuos A-} = 4 / (4+1) = 0,8 \text{ o } 80 \%$$

La sensibilidad de la prueba B fue exactamente la misma independientemente del resultado de la prueba A. De esta forma se comprueba la existencia de independencia, lo que justificaría la aplicación de ambas pruebas en paralelo, ya que:

$$\text{Sensibilidad A} = (76 + 19) / 100 = 95 \%$$

$$\text{Sensibilidad B} = (76 + 4) / 100 = 80 \%$$

$$\text{Sensibilidad A y B (paralelo)} = (76 + 19 + 4) / 100 = 99 \%$$

Ahora vamos a suponer que aplicamos, igual que en el ejemplo anterior, dos pruebas en paralelo con sensibilidades de 95 y 80 % para las pruebas A y B, respectivamente. Los resultados se presentan en el cuadro 20.

Cuadro 20. Dependencia condicional completa en la sensibilidad de dos pruebas diagnósticas aplicadas en paralelo

Condición	Prueba B+	Prueba B-	Total
Prueba A+	80	15	95
Prueba A-	0	5	5
Total	80	20	100

La sensibilidad de la prueba A es del 95 % [(80 + 15) / 100] y la de la prueba B es del 80 % (80 / 100). Ahora vamos a evaluar si ambas pruebas son independientes. De acuerdo con la definición dada anteriormente, debemos comprobar los siguientes supuestos:

$$\text{Sensibilidad Prueba B en individuos A+} = 80 / (80+15) = 0,84 \text{ o } 84 \%$$

$$\text{Sensibilidad Prueba B en individuos A-} = 0 / (0+5) = 0,0 \text{ o } 0 \%$$

Como se puede observar, la sensibilidad de B es diferente entre los individuos que dieron positivos o negativos a la prueba A. Dicho de otra manera, la sensibilidad de

B está condicionada por el resultado de la prueba A. Este caso se denomina dependencia condicional entre pruebas diagnósticas. Evidentemente, si aplicamos estas dos pruebas en paralelo no generamos ninguna ventaja adicional al usar la prueba A solamente, ya que:

$$\text{Sensibilidad Prueba A} = (80 + 15) / 100 = 95 \%$$

$$\text{Sensibilidad Prueba A y B (en paralelo)} = (80 + 15 + 0) / 100 = 95 \%$$

Igualmente podemos analizar la presencia de dependencia en la especificidad de las pruebas diagnósticas. Para ello debemos comprobar los siguientes supuestos:

- a)** La especificidad de la segunda prueba no depende de que el resultado de la primera haya sido positivo o negativo en la población sana.
- b)** La segunda prueba detecta todos los animales sanos que su especificidad le permite entre aquellos que resultaron positivos a la primera.

Para ello vamos a considerar los mismos valores del Cuadro 19.

$$\text{Especificidad Prueba B en individuos A+} = 120 / (120 + 30) = 0,80 \text{ o } 80 \%$$

$$\text{Especificidad Prueba B en individuos A-} = 680 / (680 + 170) = 0,80 \text{ o } 80 \%$$

De esta manera se comprueba la independencia en la especificidad de ambas pruebas diagnósticas aplicadas consecutivamente y se justifica su empleo combinado ya que mejora el rendimiento de cada una de las pruebas aplicadas en forma individual:

$$\text{Especificidad Pruebas A y B (consecutivas)} = (120 + 170 + 680) / 1000 = 97 \%$$

Analicemos ahora la información contenida en el Cuadro 21. La especificidad de la prueba A es del 85 %, mientras que la de la prueba B es del 80 %. Evaluemos si ambas pruebas son independientes. Según la definición dada anteriormente, debemos comprobar los siguientes supuestos:

$$\text{Especificidad Prueba B en individuos A+} = 0 / (150) = 0,0 \text{ o } 0 \%$$

$$\text{Especificidad Prueba B en individuos A-} = 800 / (800 + 50) = 0,941 \text{ o } 94,1 \%$$

Como se advierte, la especificidad de B es diferente entre los individuos que dieron positivos o negativos a la prueba A. Dicho de otro modo, la especificidad de B está condicionada por el resultado de la prueba A. Éste es otro caso de dependencia condicional entre pruebas diagnósticas. Obviamente, si aplicamos estas dos pruebas consecutivamente no generemos ninguna ventaja adicional al usar la prueba A solamente, ya que:

Especificidad Prueba A = $850 / 1000 = 85 \%$

Especificidad Prueba A y B (consecutivas) = $(50 + 800 + 0) / 100 = 85 \%$

Cuadro 21. Dependencia condicional completa en la especificidad de dos pruebas diagnósticas aplicadas consecutivamente

Condición	Prueba B+	Prueba B-	Total
Prueba A+	150	0	150
Prueba A-	50	800	850
Total	200	800	1000

7. Otras estrategias

7.1. Paneles de pruebas

En el proceso de efectuar un diagnóstico, el clínico observa si un antecedente está o no presente en la anamnesis efectuada sobre el paciente, si un signo está o no presente y/o si los resultados de las pruebas de laboratorio están fuera del «rango normal». Luego de combinar esta información, el clínico emite un juicio sobre cuál es el diagnóstico más probable de una lista de diagnósticos diferenciales que él ha formulado. El clínico se enfrenta a un desafío particular en la selección de las pruebas paraclínicas (sobre todo si ellas son caras, complicadas o potencialmente perjudiciales para el paciente) que puedan ayudar a formar y no a confundir su juicio. Si la mayoría, o la totalidad, de los resultados deben ser positivos para proveer de evidencia suficiente de la enfermedad de interés, el proceso diagnóstico será bastante específico, con pocos falsos positivos y un alto VP+. Sin embargo, la sensibilidad será baja y muchos individuos con la enfermedad no serán correctamente identificados como positivos (muchos falsos negativos). Si un resultado positivo en sólo una o unas pocas pruebas es suficiente evidencia de la enfermedad de interés, el proceso será altamente sensible, con pocos falsos negativos y un alto VP-. En este caso el VP+ será bajo, por lo que la probabilidad de que un individuo esté enfermo al ser positivo no difiere mucho de la probabilidad que manejaba el clínico antes de conocer los resultados de las pruebas (Martin y Bonnett, 1987). El clínico puede pensar que la probabilidad de efectuar el diagnóstico correcto en un paciente sea asegurada mediante el uso de pruebas múltiples. El problema es que en general los hallazgos están correlacionados entre sí y, por lo tanto, el valor del segundo resultado luego de que se tiene en cuenta el primer hallazgo se ve disminuido con respecto al valor del segundo resultado por sí mismo (Smith et al., 1986).

Cuando el clínico solicita al laboratorio que efectúe una batería de pruebas diagnósticas sobre todos los animales, o una muestra de ellos, también incluye una proporción variable de animales sanos de la patología que se quiere diagnosticar. Para simplificar

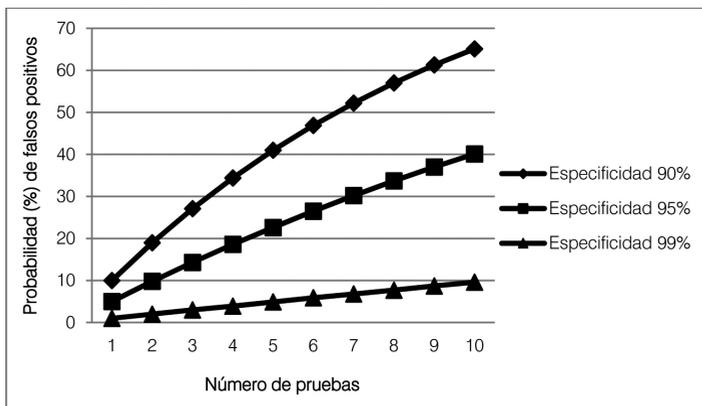
la discusión, supongamos que todas las pruebas tienen una especificidad del 80 % y son biológicamente independientes y que existen 100 animales sanos en el rodeo o la muestra enviada al laboratorio. En este caso, 20 de ellos serán dados como positivos (falsos positivos) a la primera prueba. De los 80 animales negativos a la primera prueba, otros 6,4 serán dados como positivos a la segunda prueba (falsos positivos) y así sucesivamente. Nótese que las pruebas se van efectuando sobre los que quedan negativos, de igual manera que en la estrategia de repetición. Este razonamiento es intuitivamente aplicado cuando el clínico carece de otros elementos y puede esquematizarse como:

A+ = +

A- = Realizar más pruebas.

De este modo, a medida que el número de pruebas se incrementa en un individuo, también lo hace la probabilidad de un diagnóstico falso positivo y disminuye el VP de cualquier prueba positiva (Tyler y Cullor, 1989) (Figura 1).

Figura 1. Probabilidad (%) de falsos positivos en al menos una prueba para diversos valores de especificidad y de número de pruebas no relacionadas efectuadas



La figura anterior nos debe alertar contra el uso indiscriminado de paneles, paquetes o baterías de pruebas diagnósticas en **todos** los pacientes. Este proceso no sólo es un desperdicio de material, esfuerzo y dinero sino que es improbable que mejore la habilidad del clínico en cuanto a establecer el diagnóstico correcto (Martin y Bonnett, 1987). Cuando los valores normales son definidos de acuerdo con la curva gaussiana y se efectúan varias pruebas diagnósticas en el mismo sujeto, más pruebas se realicen en un individuo sano, más alta será la probabilidad de encontrar un resultado anormal en al menos una prueba (Galen y Gambino, 1975:237). La probabilidad de que un

paciente dado esté fuera del rango normal en al menos una prueba es igual a $1 - \text{Esp.}^n$, siendo n el número de pruebas independientes efectuadas (Knapp y Clinton Miller, 1992:435). Esto llevó al siguiente aforismo: *el único paciente «normal» es aquel que aún no ha sido suficientemente investigado.*

7.2. Pruebas combinadas en tamizado

La presencia de determinados hallazgos (ej.: sombras en una radiografía), junto a otros resultados, puede llevar al diagnóstico de una enfermedad específica. En ciertas condiciones y en muy contadas ocasiones es posible incrementar simultáneamente la sensibilidad y la especificidad de este diagnóstico por medio de la utilización de pruebas múltiples (Martin, 1976). En estos casos, la asignación de los individuos al grupo que continuará en observación o al grupo que será librado de la misma dependerá del número determinado de pruebas positivas que sea requerido. La utilidad de esta estrategia está limitada por la forma de las distribuciones de la proporción de enfermos y sanos. Cuanto mayor sea la superposición de las curvas, menor será la expectativa de mejora en la discriminación de enfermos y sanos (Nissen-Meyer, 1964). Veamos este ejemplo de interpretación de tres pruebas combinadas (Riegelman y Hirsch, 1989:256):

A+ B+ = +

A+ B- = Resultado dudoso, realizar más pruebas.

A- B+ = Resultado dudoso, realizar más pruebas.

A- B- = -

Si en las muestras que arrojaron resultados clasificados como dudosos aplicamos una tercera prueba diagnóstica independiente C, obtendremos:

A+ B+ = +

A+ B- C+ = +

A- B+ C+ = +

A+ B- C- = -

A- B+ C- = -

A- B- = -

Aquí la selección de la línea divisoria entre positivos y negativos fue considerar como positivos a los individuos que arrojaron dicho resultado en dos o más pruebas. Dependiendo de la sensibilidad y especificidad de las pruebas, puede llegarse, aunque no necesariamente, a incrementar la sensibilidad y especificidad con respecto a la ejecución de una única prueba (A, B o C). Sin embargo, la sensibilidad será menor que si se

hubieran ejecutado las pruebas en paralelo, mientras que la especificidad será menor que en el caso de pruebas consecutivas, afectando consecuentemente los VP- y VP+, respectivamente.

Esta estrategia de interpretación de tres pruebas diagnósticas implica analizar cada individuo con dos pruebas y, en caso de desacuerdo en los resultados, realizar una tercera tomando como resultado aquel de la mayoría, o bien analizar cada individuo con una prueba y reanalizar los positivos con un segundo test; entonces, si hay desacuerdo, realizar un tercer test tomando como resultado final aquel de la mayoría (Riegelman y Hirsch, 1989:256). Sin embargo, aunque en algunas circunstancias el uso de pruebas combinadas pueda incrementar en algo la sensibilidad y la especificidad del diagnóstico a través de la elección de una apropiada línea de división entre el número de pruebas requerido para ser considerado positivo o negativo, siempre coexistirá el conflicto entre ambos objetivos (Nissen-Meyer, 1964).

Bibliografía

- Dohoo, I.R., Wright, P.F., Ruckerbauer, G.M., Samagh, B.S., Robertson, F.J., Forbes, L.B. (1986).** A comparison of five serological tests for bovine brucellosis. *Can. J. Vet. Res.* **50** : 485–493.
- Francis, J., Seiler, R.J., Wilkie, I.W., O’Bye, D., Lumsden, M.J., Frost, A.J. (1978).** The sensitivity and specificity of various tuberculin tests using bovine PPD and other tuberculins. *Vet. Rec.* **103** : 420–435.
- Galen, R.S., Gambino, S.R. (1975).** Beyond Normality, the Predictive Value and Efficiency of Medical Diagnoses. John Wiley & Sons, New York, 237 pp.
- Gardner, I.A., Greiner, M. (1999).** Advanced Methods for test Validation and Interpretation in Veterinary Medicine Freie. Univ. Berlin, 74 págs.
- Gutiérrez, M., Tellechea, J., García Marín, J.F. (1998).** Evaluation of cellular and serological test for the detection of *Mycobacterium bovis*-infected goats. *Vet. Microbiol.* **62**: 281–290.
- Knapp, R.G., Clinton Miller, M. (1992).** Clinical Epidemiology and Biostatistics. Natl. Med. Series, Williams & Wilkins, Harwal Publ. Co., Pennsylvania, 435 pp.
- Lightbody, K.A., Skuce, R.A., Neill, S.D., Pollock, J.M. (1998).** Mycobacterial antigen-specific antibody responses in bovine tuberculosis: an ELISA with potential to confirm disease status. *Vet. Rec.* **142**: 295–300.
- Martin, S.W. (1976).** The evaluation of tests. *Can. J. Comp. Med.* **41**: 19–25.
- Martin, S.W., Bonnett. (1987).** Clinical Epidemiology. *Can. Vet. J.* **6**: 318–325.
- Nissen-Meyer, S. (1964).** Evaluation of screening tests in medical diagnosis. *Biometrics* **20**: 730–755.
- Riegelman, R.K., Hirsch, R.P. (1989).** Cómo estudiar un estudio y probar una prueba: Lectura crítica de la literatura médica. Little Brown & Co., 2º Ed., OPS Pub. Cien. 532, 256 pp.
- Sanson, R.L. (1988).** Tuberculosis in goats. *Surveillance* **15**: 7–8.
- Smith, D.F., Munson, L., Erb, H.N. (1986).** Predictive values for clinical signs of abomasal ulcer disease in dairy cattle. *Prev. Vet. Med.* **3**: 573–580.
- Smith, R.D. (1991).** Veterinary Clinical Epidemiology. A problem-oriented approach. Butterworth-Heinemann, Reed Pub. Inc. USA, 225 pp.
- Tarabla, H.D. (2000).** Interpretación de resultados de pruebas diagnósticas en programas de control y erradicación de enfermedades. XXI Congreso Mundial Buiatría, Punta del Este, Uruguay.
- Tyler, J.W., Cullor, J.S. (1989).** Titers, tests, and truisms: Rational interpretation of diagnostic serologic testing. *J. Am. Vet. Med. Assoc.* **194**: 1550–1558.

Capítulo 11

Poder discriminatorio de la prueba diagnóstica y prevalencia de la enfermedad

En general, las pruebas usadas en estudios poblacionales fueron desarrolladas y utilizadas primero como pruebas diagnósticas en individuos. En el capítulo anterior vimos ejemplos de esto último, mientras que en éste se abordarán aspectos relacionados con la utilización de los test diagnósticos en estudios epidemiológicos. Para el clínico, el objetivo del diagnóstico es clasificar al paciente en un grupo adecuadamente definido para que las acciones subsecuentes (ya sean médicas, quirúrgicas u otras) optimicen la salud del paciente. Cuando el propósito de la prueba es diagnóstico, el interés fundamental es la evaluación de la exactitud de la prueba en cuanto a clasificar correctamente animales sanos y enfermos a través de la especificidad y sensibilidad de la prueba y a determinar la probabilidad de que un animal positivo esté realmente enfermo (VP+), uno sano esté en realidad sano (VP-) y sus situaciones complementarias (probabilidad de falsos positivos o negativos).

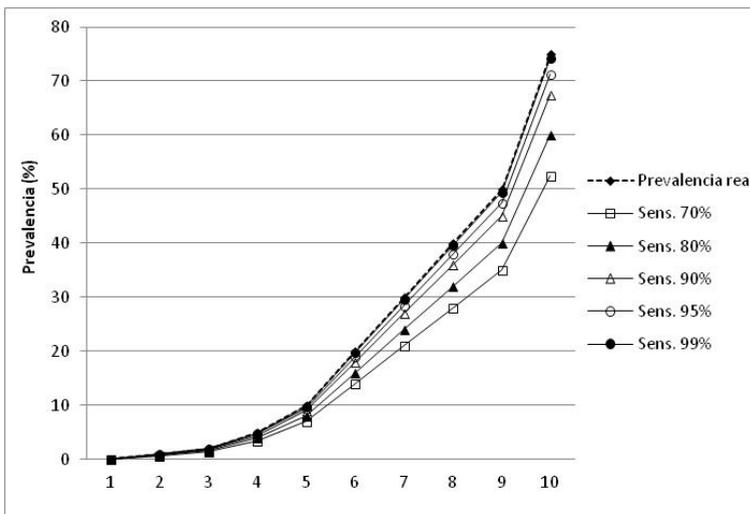
El uso de pruebas diagnósticas en estudios epidemiológicos requiere que el investigador tenga un concepto claro del objetivo de su investigación, que adopte un diseño adecuado del relevamiento, que maneje una definición precisa tanto de la población a la cual se quiere inferir los resultados como de la muestra que va a ser usada para representar dicha población y, obvio es decirlo, que posea un profundo conocimiento de las virtudes y limitaciones de la prueba a utilizar. En la estimación de la prevalencia de una enfermedad, la preocupación principal no es la exactitud del diagnóstico individual, sino el sesgo y la precisión con los cuales se puede estimar el parámetro poblacional (ej.: la prevalencia real de la enfermedad) (McClish y Quade, 1985). Aquí la prueba se utiliza ya no para tomar decisiones terapéuticas, quirúrgicas o de otro tipo en un paciente en particular, sino para cuantificar la ocurrencia de enfermedades en poblaciones, generalmente en forma de estudios de prevalencia. La elección de la prueba diagnóstica va a influenciar incluso la determinación de la cantidad de animales a muestrear (Farver, 1984). Los diagnósticos falsos positivos en estos casos

no tienen una consecuencia directa sobre los individuos que componen la población, sino en la sobre o subestimación de la cantidad de enfermos presentes en esa población (Marchevsky, 1974).

1. Efecto de la sensibilidad de la prueba diagnóstica sobre la prevalencia aparente de la enfermedad

En general, las variaciones moderadas en la sensibilidad de la prueba no afectan mayormente las estimaciones de prevalencia. Como se observa en la Figura 1, donde se asume una especificidad constante del 100 %, sólo si se utilizan pruebas con baja sensibilidad y en casos de prevalencias reales muy altas puede existir un efecto de subestimación de la misma. En diagnósticos individuales se necesitan pruebas altamente sensibles, pero en estudios poblacionales esto no es tan crítico, dado que una prueba con un nivel de sensibilidad aceptable puede ser igualmente satisfactoria. Esto es particularmente cierto en poblaciones con una baja proporción de individuos afectados.

Figura 1. Prevalencia aparente de la enfermedad en distintos valores de prevalencia real y de sensibilidad de la prueba (especificidad = 100 %).

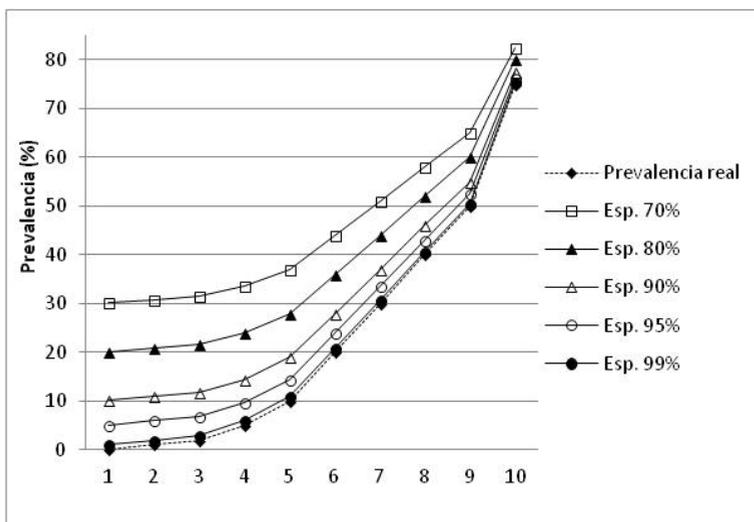


2. Efecto de la especificidad de la prueba diagnóstica sobre la prevalencia aparente de la enfermedad

Las estimaciones de la prevalencia de una enfermedad se ven muy influenciadas por variaciones en la especificidad de la prueba. Cuando se analicen los resultados de un relevamiento en una población se debe prestar particular atención al error provo-

cado fundamentalmente por la falta de especificidad para no sobreestimar la prevalencia de la enfermedad. Este efecto se hace más marcado a medida que la prevalencia de la enfermedad es más baja (Figura 2). La asignación a una enfermedad de una prevalencia mayor a la real puede adquirir gran importancia, particularmente de índole económica. Por otra parte, en el diseño del relevamiento, la especificidad juega un rol relevante en la determinación del tamaño de la muestra (Farver, 1984).

Figura 2. Prevalencia aparente de la enfermedad en distintos valores de prevalencia real y de especificidad de la prueba (sensibilidad = 100 %).



Como se vio en el ejemplo de la tuberculosis bovina en el punto «Prevalencia real y prevalencia aparente. Un ejemplo», la sobreestimación de la prevalencia de una enfermedad puede ser notable, particularmente en prevalencias reales bajas. En el contexto de un programa de erradicación de la enfermedad significa que se subestimarían groseramente los progresos del programa (Tarabla, 1995).

3. Efecto de la sensibilidad y especificidad de la prueba diagnóstica y la prevalencia real de la enfermedad sobre la prevalencia aparente.

Un ejemplo

3.1. Cambios en la especificidad

Supongamos que tenemos una enfermedad con una prevalencia real del 1 % y una prueba con una sensibilidad del 95 % y una especificidad del 95 %. Aplicando el Teorema de Bayes tendremos que:

$$P(E+ | T+) = \frac{0,01 \times 0,95}{0,01 \times 0,95 + (1 - 0,01) \times (1 - 0,95)} = \frac{0,0095}{0,059} = \mathbf{0,161}$$

Esto, llevado a porcentaje, significa un VP+ del 16,1 %, es decir que, de cada 100 reacciones positivas, sólo 16 individuos están realmente enfermos.

Este ejemplo se puede ver claramente en el Cuadro 1, donde se muestra que la prevalencia aparente es seis veces mayor que la real:

Cuadro 1. Prevalencia aparente cuando la prevalencia real es del 1 % y la sensibilidad y especificidad de la prueba son del 95 %

Condición	Enfermos	Sanos	Total
Test positivos	95	495	590
Test negativos	5	9405	9410
Total	100	9900	10 000

$$\text{Prevalencia aparente} = \frac{590}{10\,000} \times 100 = \mathbf{5,9\%}$$

Supongamos ahora que se emplee una prueba menos específica (90 %). Aplicando el Teorema de Bayes obtendremos que el VP+ cae casi a la mitad (8,8 %), es decir que, de cada 100 reacciones positivas, sólo 9 individuos están realmente enfermos. Por otra parte, la prevalencia aparente se ve fuertemente influenciada y es ahora casi 11 veces mayor que la real (Cuadro 2):

Cuadro 2. Prevalencia aparente cuando la prevalencia real es del 1 %, la sensibilidad del 95% y la especificidad del 90 %

Condición	Enfermos	Sanos	Total
Test positivos	95	990	1085
Test negativos	5	8910	8915
Total	100	9900	10 000

$$\text{Prevalencia aparente} = \frac{1085}{10\,000} \times 100 = \mathbf{10,85\%}$$

Si repetimos el ejercicio anterior con una enfermedad de mayor prevalencia (prevalencia real del 30 %, prueba con una sensibilidad y una especificidad del 95 %) y aplicando el Teorema de Bayes, tendremos un VP+ del 89,1 %. Esto indica que 9 de cada 10 reacciones positivas corresponden a individuos realmente enfermos. En tanto, la prevalencia aparente es muy similar a la real (Cuadro 3):

Cuadro 3. Prevalencia aparente cuando la prevalencia real es del 30 % y la sensibilidad y especificidad de la prueba es del 95 %

Condición	Enfermos	Sanos	Total
Test positivos	2850	350	3200
Test negativos	150	6650	6800
Total	3000	7000	10 000

$$\text{Prevalencia aparente} = \frac{3200}{10\,000} \times 100 = \mathbf{32\%}$$

Supongamos ahora que utilizamos una prueba con una especificidad del 90 %. Aplicando el Teorema de Bayes obtendremos que el VP+ = 80,3 %. Nótese que el VP+ sólo descendió un 10 % y no a la mitad como cuando la enfermedad tenía una prevalencia del 1 %. Por lo tanto, 8 de cada 10 reacciones positivas pertenecen a individuos realmente enfermos y la prevalencia aparente se ve menos afectada que en el ejemplo del 1 % (Cuadro 4):

Cuadro 4. Prevalencia aparente cuando la prevalencia real es del 30 %, la sensibilidad del 95 % y la especificidad del 90 %

Condición	Enfermos	Sanos	Total
Test positivos	2850	700	3550
Test negativos	150	6300	6450
Total	3000	7000	10 000

$$\text{Prevalencia aparente} = \frac{3550}{10\,000} \times 100 = \mathbf{35,5\%}$$

3.2. Cambios en la sensibilidad

Tomando el mismo ejemplo de una enfermedad con una prevalencia real del 1 % y una prueba con una sensibilidad y especificidad del 95 %, vimos que el VP+ es del 16,1 % y la PA del 5,9 %.

Supongamos ahora que disminuya la sensibilidad al 90 % mientras que la especificidad y la PR permanecen constantes. Si repetimos los cálculos anteriores, obtendremos un VP+ del 15,4 % y una PA del 5,9 %. Esta disminución del 5 % en la sensibilidad no influyó mayormente VP+ ni la PA.

Con una enfermedad de mayor prevalencia (prevalencia real del 30 %, y una prueba con una especificidad y sensibilidad del 95 %) el VP+ es del 89,16 % y la PA del 32 %.

Si se disminuye la sensibilidad al 90 %, mientras que la especificidad y la PR permanecen constantes, el VP+ llega al 88,5 %. Nótese cómo con un descenso del 5 % en la sensibilidad se produjo una disminución muy pequeña en el VP+ mientras que la prevalencia aparente fue muy similar (30,50 %).

4. Estimación de la prevalencia real de la enfermedad

La prevalencia real de una enfermedad puede ser estimada a partir de su prevalencia aparente y la sensibilidad y especificidad de la prueba diagnóstica empleada aplicando la siguiente fórmula desarrollada por Marchevsky (1974).

$$\text{Prevalencia Real Estimada} = \frac{\text{PA} - (1 - \text{Esp.})}{1 - [(1 - \text{Esp.}) + (1 - \text{Sens.})]}$$

Aplicándola al ejemplo del Cuadro 4 en el cual la prevalencia aparente era del 35,5 %, la sensibilidad del 95 % y la especificidad del 90 %, obtendremos:

$$\text{Prevalencia Real Estimada} = \frac{0,355 - (1 - 0,90)}{1 - [(1 - 0,90) + (1 - 0,95)]} = 0,30 \text{ o } 30 \%$$

De igual manera, si la aplicamos en el ejemplo del Cuadro 3 del Capítulo 7 (intradermorreacción en pliegue anocaudal, sensibilidad del 81,8 % y especificidad del 96,3 % (Francis et al., 1978), con una prevalencia aparente del 4,481 %):

$$\text{Prevalencia Real Estimada} = \frac{0,04481 - (1 - 0,963)}{1 - [(1 - 0,963) + (1 - 0,818)]} = 0,01 \text{ o } 1 \%$$

La estimación de la prevalencia real se hace menos exacta a medida que la prueba es menos sensible y menos específica. Cabe hacer notar que cuando la sensibilidad y especificidad de la prueba son complementarias (es decir que, sumadas dan el 100 %; por ejemplo, una prueba que sea 60 % sensible y 40 % específica), este método no puede ser utilizado para estimar la prevalencia real a partir de la aparente. Afortunadamente esta situación es altamente improbable.

Otra forma de resolver el problema de la estimación de la prevalencia real a partir de datos de prevalencia aparente (PA) y las características intrínsecas de las pruebas diagnósticas (sensibilidad y especificidad) es lo que se conoce como estimador de Rogan-Gladen (P_{RG}) (Vázquez, 1990). Si se denomina P a la proporción de animales que ha sido clasificadas como enfermas en la población y PR a la prevalencia real, se observa que P tiene dos componentes: uno que procede de los animales enfermos que han sido clasificados como tales (verdaderos positivos) y otro que corresponde a los animales que no teniendo la enfermedad, fueron erróneamente clasificados

como tales (falsos positivos). La proporción de animales que son clasificados como enfermos es, entonces:

$$PA = PR \times Se + (1 - PR) \times (1 - Esp)$$

\downarrow
 Verdaderos positivos

\downarrow
 Falsos positivos

Como se puede observar, la prevalencia aparente depende tanto de la prevalencia real como de la sensibilidad y de la especificidad de la prueba diagnóstica empleada. Despejando PR en la ecuación anterior, se obtiene una estimación de la prevalencia (P_{RG}) que puede ser empleada en aquellas situaciones donde la sensibilidad y la especificidad de la prueba diagnóstica son conocidas o pueden ser estimadas:

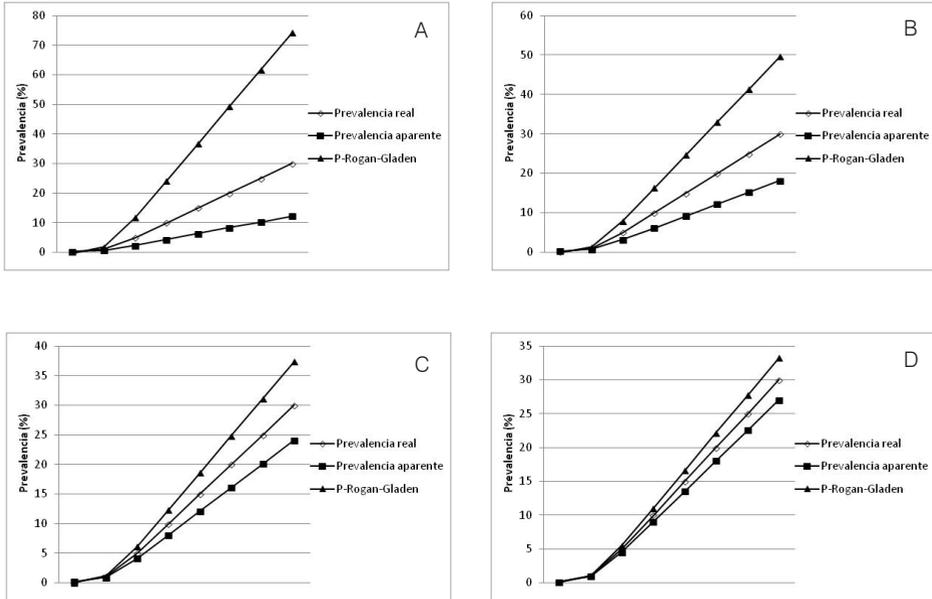
$$P_{RG} = \frac{PA + Esp - 1}{Sens. + Esp. - 1}$$

Esta fórmula ofrece muy buenos ajustes a la prevalencia real de la enfermedad cuando los valores de sensibilidad y especificidad son superiores al 80 %. Al igual que en el caso anterior, esta aproximación no es válida cuando la sumatoria de la sensibilidad y la especificidad de la prueba diagnóstica es igual a 1.

Otra consideración que debe tenerse en cuenta para el uso del estimador de Rogan–Gladen es que en casos con prevalencias reales muy bajas o muy altas pueden generarse valores menores a 0 o mayores a 1, respectivamente. Entonces, es necesario corregir la fórmula y delimitar sus resultados en valores mínimos y máximos de 0 y 1, respectivamente.

En la Figura 3 se presenta el grado de ajuste de la prevalencia real empleando ambas fórmulas, la de Marchevsky y la de Rogan–Gladen. En la figura sólo se expone el ajuste con cuatro combinaciones de sensibilidad y especificidad de la prueba diagnóstica.

Figura 3. Ajuste de la prevalencia real según las fórmulas de Marchevsky (prevalencia aparente) y Rogan–Gladen para diferentes combinaciones de sensibilidad y especificidad.



Nota: Valores de Sensibilidad = Especificidad, A) 70 %, B) 80 %, C) 90 % y D) 95 %.

En términos generales, el ajuste de ambas fórmulas es mejor a medida que la prevalencia real de la enfermedad disminuye. Independientemente de la prevalencia real y de la sensibilidad o especificidad de la técnica diagnóstica, el estimador de Rogan–Gladen tiende a sobrestimar la prevalencia real, mientras que la fórmula de Marchevsky la subestima. Obviamente, a medida que la sensibilidad y especificidad de la prueba diagnóstica se incrementan, el ajuste de la prevalencia real mejora.

Cuando la sensibilidad y especificidad de la prueba diagnóstica son bajas, tal como se muestra en las Figuras 3 A y B, el error en la estimación mediante el estimador de Rogan–Gladen tiende a ser muy grande. Si bien lo anterior también aplica para el caso de la fórmula de Marchevsky, el error en el ajuste no está tan influenciado por las características propias de la prueba diagnóstica.

Bibliografía

- Farver, T.B. (1984).** Some practical considerations in sampling livestock populations to estimate prevalence and other parameters. *Prev. Vet. Med.* **2**: 453–462.
- Francis, J., Seiler, R.J., Wilkie, I.W., O’Byrne, D., Lumsden, M.J., Frost, A.J. (1978).** The sensitivity and specificity of various tuberculin tests using bovine PPD and other tuberculins. *Vet. Rec.* **103**: 420–435.
- Marchevsky, N. (1974).** Errores en las estimaciones de prevalencia en estudios de población: Un método para calcular la prevalencia real. *Zoonosis.* **16**: 81–109.
- McClish, D., Quade, D. (1985).** Improving estimates of prevalence by repeated testing. *Biometrics.* **41**: 81–89.
- Tarabla, H.D. (1995).** Efecto del poder discriminatorio de la prueba diagnóstica sobre la estimación de la prevalencia de la enfermedad. Resúmenes 11 Congr. Arg. y 11 Congr. Latinoam. Zoonosis, Buenos Aires, Argentina, **p.** 209.
- Vázquez Z.G. (1990).** Epidemiología veterinaria y salud animal. Ed. Limuds, SA de CV México DF (México).

Capítulo 12

Interpretación de pruebas diagnósticas a nivel de grupos

En algunas ocasiones los veterinarios estamos más interesados en determinar la condición sanitaria de un grupo de animales que en el resultado individual de la prueba diagnóstica (Christensen y Gardner, 2000).

Una agrupación de animales, ya sea una piara, un galpón de gallinas, un rodeo de bovinos o un tanque de peces, puede ser clasificado como sana o enferma basado en:

- a)** los resultados obtenidos en varias muestras individuales tomadas de todos o de al menos 2 de los individuos que componen el grupo (Gardner y Greiner, 1999:74),
- b)** una sola muestra compuesta (MC) de todos los animales que lo componen (ej.: conteos de células somáticas superiores a los 500 000/ml en la leche total del tambo pueden señalar rodeos con problemas de mastitis subclínica).

En el primer caso, un rodeo puede ser clasificado como enfermo si por lo menos un animal dio positivo (ej.: inmunofluorescencia directa en campilobacteriosis genital bovina) o si la proporción de animales positivos supera un límite preestablecido (ej.: proporción de animales con reacción a la Prueba de California superior al grado 2 en mastitis bovina).

En cambio, cuando la prueba diagnóstica se aplica a una sola muestra que contiene todos los individuos que integran el grupo, la interpretación del resultado será similar a la que se realiza con las muestras individuales.

Los casos más comunes en los que se emplean estas aproximaciones son cuando se desea, por ejemplo, certificar a una piara como libre de una enfermedad, analizar la eficacia de los programas de control o erradicación, estudiar los factores de riesgo o evaluar el riesgo de introducir una enfermedad a una región (Houe et al., 2003).

Aunque los grupos son variados en función de la especie, es relativamente frecuente, aunque erróneo, el empleo del término «rodeo» para definir la agrupación de animales de cualquier especie.

La interpretación de los resultados de una prueba diagnóstica a nivel grupal resulta algo más complicada que a nivel individual dados los múltiples factores que influyen sobre el grupo. Cuando se aplica una prueba con fines diagnósticos en un rodeo se deben considerar dos valores umbrales: 1) el valor de corte seleccionado para la prueba diagnóstica individual y 2) el número (o proporción) de animales con resultados por sobre el valor umbral de la prueba que son requeridos para que el grupo sea considerado libre de la enfermedad que se analiza.

1. Sensibilidad y especificidad a nivel grupal

La interpretación de la sensibilidad y especificidad a nivel grupal tiene la misma lógica que cuando se aplican estos conceptos de manera individual. De esta forma, la sensibilidad a nivel de grupo (Sens.^g) es la habilidad de una prueba diagnóstica para clasificar como positivos a los grupos verdaderamente afectados por una enfermedad en particular. La especificidad a nivel de grupo (Esp.^g) es la habilidad de una prueba diagnóstica para identificar como negativos a aquellos grupos que están libres de la enfermedad. Cada una de ellas depende de: la sensibilidad y especificidad de la prueba individual, el tamaño de la muestra (o del grupo en caso de muestrear todos los animales que lo componen), el límite por sobre el cual un grupo es declarado positivo, la distribución de la enfermedad en los grupos muestreados y la posibilidad de correlación entre los resultados dentro de cada grupo.

Una vez que un grupo de animales es clasificado como no infectado, todos los animales que lo componen asumen automáticamente el mismo estatus. No obstante, cuando el grupo es considerado como infectado, los animales individualmente no adoptan esta condición puesto que algunos de ellos pueden no poseer la enfermedad. Esto ocurre cuando los individuos están en una etapa inicial de la enfermedad o por la susceptibilidad/resistencia propia de los hospedadores (edad, sexo o estado inmunitario) (Christensen y Gardner, 2000).

Diferentes factores pueden afectar tanto la sensibilidad como la especificidad a nivel de grupos, entre los que podemos mencionar los siguientes:

- a)** Sensibilidad y especificidad de la prueba diagnóstica a nivel individual (si la prueba es muy sensible no tendrá necesidad de muestrear muchos animales).
- b)** Número de animales dentro del grupo que serán analizados. En este capítulo se asumirá que el número de animales a muestrear es pequeño con relación al tamaño total del grupo, por lo cual se empleará la distribución binomial para todos los cálculos que se realicen. Si se muestrea una gran proporción de los animales que componen el grupo, se debería emplear la aproximación hipergeométrica para efectuar los cálculos.
- c)** Prevalencia de la enfermedad dentro del grupo (si es alta, una pequeña muestra probablemente contendrá al menos un animal enfermo).
- d)** Número de animales que deberán ser positivos a la prueba diagnóstica para que el grupo sea considerado positivo.

e) Variabilidad de la enfermedad en los animales que componen el grupo (algunos pueden estar altamente infectados y otros no).

f) Correlación entre los resultados de diferentes animales dentro del grupo. Independientemente de su sensibilidad y especificidad, si una prueba en un animal es positiva, es probable que igual prueba en otro animal del mismo grupo también sea positiva, aun si los dos animales son libres de la enfermedad (ej.: en tuberculosis bovina, la exposición natural a otros *Mycobacterium* que producen reacciones cruzadas puede ocurrir con más frecuencia en algunos grupos que en otros).

La prevalencia aparente (PA) de una enfermedad determina la probabilidad de que un animal sea positivo a la prueba diagnóstica. Por eso, la probabilidad de que un animal dé un resultado negativo a la prueba es igual a $1 - PA$. Asumiendo que la sensibilidad y especificidad de la prueba son constantes para todos los animales del grupo, podemos estimar la probabilidad de no encontrar animales positivos en el mismo $P(T-)$:

$$P(T-) = (1 - PA)^n$$

Por lo tanto, la probabilidad de que al menos un animal dé positivo a la prueba diagnóstica (Sens.^g) será igual a $1 - (1 - PA)^n$.

La Esp.^g es la probabilidad de que n animales de la muestra en un grupo verdaderamente negativo den un resultado negativo. Como la probabilidad de que cada animal verdaderamente negativo dé un resultado negativo es la especificidad de la prueba diagnóstica a nivel individual, la fórmula para calcular la Esp.^g es:

$$\text{Especificidad a nivel de grupo} = \text{Esp.}^n$$

A partir de la fórmula anterior podemos estimar la probabilidad de que al menos un animal sea detectado positivo (falso positivo) cuando la enfermedad está ausente del grupo [$P(T+ | E-)$]:

$$P(T+ | E-) = 1 - \text{Esp.}^n$$

Estas relaciones se observan en el Cuadro 1:

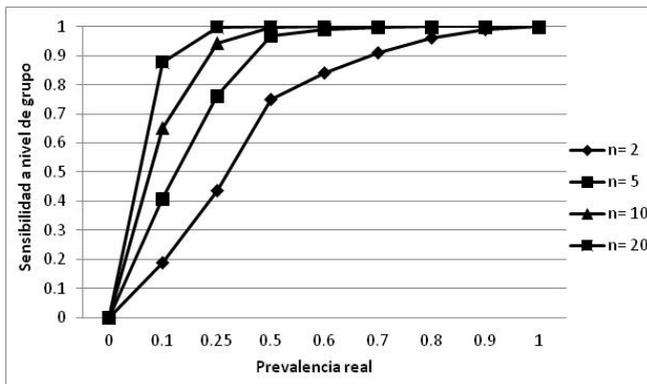
Cuadro 1. Expresión probabilística de la interpretación de las pruebas diagnósticas a nivel de grupo (Valor umbral ≥ 1 animal positivo)

Condición	Rodeo infectado	Rodeo no infectado
Al menos un animal positivo (grupo positivo)	$1 - (1 - PA)^n$	$1 - (Esp)^n$
Ningún animal positivo (grupo negativo)	$(1 - PA)^n$	$Esp.^n$
	1	1

De estas fórmulas se desprende que la sensibilidad a nivel de grupo aumenta cuando:

- a) se incrementa el número de animales a muestrear. De este modo, la probabilidad de realizar diagnósticos falsos negativos a nivel de grupo disminuye a medida que aumentamos el número de animales a muestrear (Figura 1);
- b) aumenta la prevalencia de la enfermedad en el grupo (Figura 1);
- c) se reduce la especificidad de la prueba diagnóstica;
- d) se reduce el número de animales positivos necesarios (valor umbral) para clasificar a un grupo de animales como infectado.

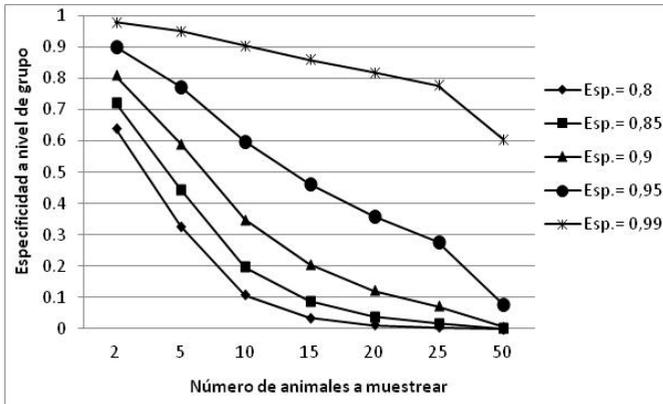
Figura 1. Variación de la Sens.^g en función del número de animales a muestrear y la prevalencia real de la enfermedad



Por otro lado, la Esp.^g no se modifica sustancialmente con los cambios en la prevalencia real de la enfermedad mientras que se incrementa cuando:

- a) aumenta la especificidad de la prueba diagnóstica (Figura 2);
- b) disminuye el número de animales muestreados (Figura 2);
- c) aumenta el número de animales positivos necesarios (valor umbral) para clasificar a un grupo como infectado.

Figura 2. Variación de la Esp.^g en función del número de animales a muestrear y la especificidad de la prueba diagnóstica



1.1. Valor umbral

Hasta ahora hemos empleado un valor umbral de al menos un animal positivo dentro del grupo para que el mismo sea considerado infectado. Obviamente podemos emplear cualquier otro valor umbral, solo que las fórmulas para calcular la sensibilidad y especificidad a nivel de grupo se complican ligeramente. Para hacerlo se deben generar expresiones binomiales como se muestran a continuación:

$$\text{Especificidad a nivel de grupo} = \sum_{i=0}^{c-1} \frac{n!}{i! \times (n-i)!} * p_f^i * q_f^{n-i}$$

$$\text{Sensibilidad a nivel de grupo} = \sum_{i=c}^n \frac{n!}{i! \times (n-i)!} * p_d^i * q_d^{n-i}$$

donde:

n= tamaño de la muestra

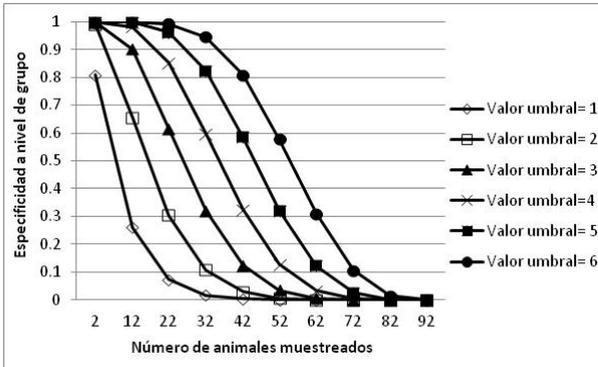
c= Número crítico o valor umbral (número de animales positivos para que el grupo sea clasificado como infectado)

p_f = PA si el grupo es verdaderamente libre

p_d = PA si el grupo es verdaderamente positivo

Consideremos una prueba que se aplica a nivel de grupo (n= 100) para clasificarlo como infectado o no infectado. La prueba tiene una sensibilidad y especificidad del 90 %. El grupo tiene una prevalencia real del 10 % y se desea estimar la especificidad de la prueba a nivel rodeo en función del número de animales a muestrear y el número de animales que deben dar positivo para que el grupo sea considerado como infectado (Figura 3).

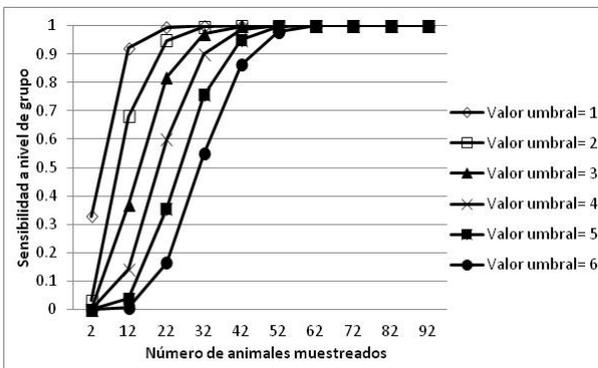
Figura 3. Variación en la Esp.^g (n= 100) en función del número de animales muestreados y el número de animales positivos necesarios para considerar al grupo como infectado (valor umbral).



Como se observa en la Figura 3, para cada valor umbral la especificidad se reduce a medida que aumentamos el número de animales a muestrear. Por otro lado, a medida que aumentamos el número de animales que deben dar positivo a la prueba diagnóstica para que el grupo sea considerado como infectado, la especificidad a nivel de grupo también se incrementa.

Por el contrario, la sensibilidad a nivel de grupo aumenta a medida que la proporción de animales muestreados en la población es mayor. Si consideramos a un grupo de animales como infectado cuando al menos uno de ellos resulta positivo a la prueba diagnóstica, muestreando aproximadamente el 10 % del grupo estaremos logrando una sensibilidad superior al 90 %. A medida que aumentamos el valor umbral, la sensibilidad a nivel de grupo se ve disminuida y, para compensar esa pérdida de sensibilidad, se debería muestrear una mayor proporción de animales (Figura 4).

Figura 4. Variación en la Sens.^g (n= 100) en función del número de animales muestreados y el número de animales positivos necesarios para considerar al grupo como infectado (valor umbral)



1.2. Valores predictivos a nivel grupal

La interpretación de los valores predictivos a nivel de grupo no difiere de su acepción individual. Todos los factores que afectan la sensibilidad y especificidad a nivel grupal, también impactarán en los valores predictivos. De esta forma, los valores predictivos dependerán de la sensibilidad y especificidad a nivel de grupo y de la prevalencia real de la enfermedad (PR). De esta forma tenemos que:

$$\text{Valor predictivo positivo a nivel de grupo (VP}^{+g}\text{)} = \frac{\text{Sens.}^g * \text{PR}}{\text{Sens.}^g * \text{PR} + (1 - \text{PR}) * (1 - \text{Esp.}^g)}$$

$$\text{Valor predictivo negativo a nivel de grupo (VP}^{-g}\text{)} = \frac{\text{Esp.}^g * (1 - \text{PR})}{\text{Esp.}^g * (1 - \text{PR}) + \text{PR} * (1 - \text{Sens.}^g)}$$

En la mayor parte de los casos el veterinario se encuentra en situaciones en las cuales desconoce los valores de sensibilidad y especificidad a nivel de grupo, al igual que los valores predictivos. No obstante, se ve obligado a interpretar resultados a nivel grupal sin esa información. En aquellas situaciones donde la prevalencia real de la enfermedad es alta, sacar conclusiones sobre el estatus del grupo puede resultar sencillo, pero cuando la prevalencia no es muy elevada (aproximadamente menor al 20 %) emitir conclusiones puede ser más difícil. Esto se debe a que, al igual que en la interpretación del VP+ a nivel individual, el VP+^g está relacionado directamente con la prevalencia real en el grupo, por lo que, a medida que ésta disminuye, también lo hará la probabilidad de que el grupo esté verdaderamente infectado cuando la prueba grupal arroje un resultado positivo.

1.3. Cálculo del número de animales a muestrear

Un aspecto crítico para catalogar un grupo con base en el muestreo de animales que lo componen es determinar el número apropiado de animales a muestrear (n). Para estimar el n de la muestra se necesitan dos valores: 1) el nivel de confianza (C) o nivel de Sens.^g (normalmente no inferior a 0,95) y 2) la prevalencia real de la enfermedad en el grupo (PR). Se puede emplear la siguiente fórmula (Gardner y Greiner, 1999:74):

$$n \geq \frac{\log(1 - C)}{\log(1 - \text{PR})}$$

Si se desea incluir en el cálculo la sensibilidad y especificidad de la prueba diagnóstica empleada (Gardner y Greiner, 1999:74):

$$n \geq \frac{\log(1 - C)}{\log(\text{Esp.} * (1 - \text{PR}) + (1 - \text{Sens.}) * \text{PR})}$$

2. Análisis de muestras compuestas

A veces, para reducir los costos o cuando no son necesarios los resultados individuales o no están disponibles, se puede considerar la toma de muestras correspondientes a varios animales como una sola muestra. De esta forma, es posible tomar muestras compuestas (MC) de heces, leche, huevos o tejidos animales según la enfermedad a evaluar.

La principal ventaja de esta estrategia en comparación con el muestreo individual es que hay más individuos representados, sin que esto genere mayores costos de laboratorio. Así, al incrementarse el número de animales muestreados, se aumenta la Sens.^g (y como contrapartida se reduce la Esp.^g). La mayor desventaja de esta estrategia de muestreo es que se puede ver reducida la sensibilidad analítica de la prueba diagnóstica dado el efecto de dilución que puede producirse cuando pocas muestras positivas son mezcladas con muchas muestras negativas (Christensen y Gardner, 2000).

La sensibilidad de la MC (Sens.^{MC}) es la probabilidad de que una muestra resulte positiva puesto que la enfermedad está presente en la MC [P(MC T+ | MC E+)] y la especificidad de la muestra tomada en esta forma (Esp.^{MC}) es la probabilidad de que una muestra arroje un resultado negativo cuando la enfermedad no está presente en la MC [P(MC T- | MC E-)].

Esta aproximación es más eficiente cuando la prevalencia de la enfermedad en la población es baja. Si el número de animales en la MC (*m*) es moderadamente alto, la sensibilidad del test basado en esa muestra (Sens.^{MC}) es menor que la sensibilidad de la prueba a nivel individual. La sensibilidad a nivel de grupo basado en *r* MC, cada una conteniendo material de *m* animales, es igual a (Dohoo et al., 2003:706).

$$\text{Sens.}^g = 1 - [(1 - (1 - \text{PR})) * (1 - \text{Sens.}) + (1 - \text{PR})^m * \text{Esp.}^{\text{MC}}]^r$$

Si el grupo de animales no presenta la enfermedad, la especificidad a nivel grupal basada en una MC es (Esp.^{MC})^r y, si no hay agrupamientos entre las MC, entonces Esp.^{MC} = Esp.^m. Si se realiza un muestreo de este tipo en un grupo de animales que se considera libre de la enfermedad, la prevalencia aparente en el grupo = 1 - Esp.^g = 1 - (Esp.^{MC})^r, que nos permite resolver la Esp.^{MC} cuando es desconocida.

$$\text{Esp.}^g = (\text{Esp.}^{\text{MC}})^r = (\text{Esp.}^m)^r$$

Tomemos como ejemplo un caso donde se quiere determinar la presencia de una enfermedad a través de la detección de anticuerpos en leche a través de la toma de MC. La técnica empleada tiene una sensibilidad=0,647 y una especificidad=0,981. Se toman 6 MC de leche (r) y cada una de ellas está constituida por 5 animales (m).

Si el grupo es libre de la enfermedad, la especificidad a nivel grupal será:

$$\text{Esp.}^g = (\text{Esp.}^{\text{MC}})^r = (\text{Esp.}^m)^r = (0,981^5)^6 = 0,524$$

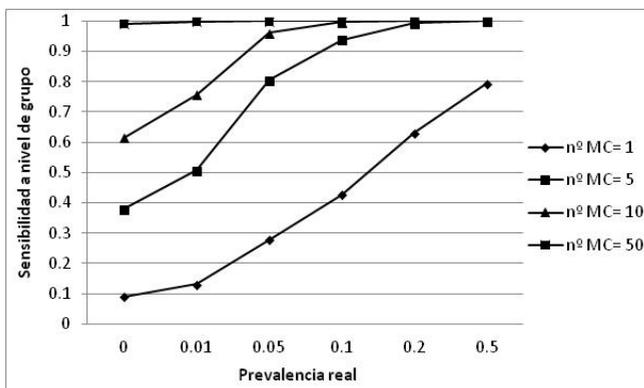
Si el grupo tuviese una prevalencia real del 12 %, asumiendo que no hay efectos de dilución, entonces la sensibilidad a nivel grupal sería:

$$\text{Sens.}^g = 1 - [(1 - (1 - 0,12)) \times (1 - 0,647) + (0,88)^5 \times 0,909]^6$$

$$\text{Sens.}^g = 0,755$$

La sensibilidad a nivel de grupo se incrementa cuando se muestrean más animales a través del uso de MC. Si bien en este caso la sensibilidad a nivel grupal calculada se incrementa a medida que aumenta la prevalencia real de la enfermedad, el mayor efecto se observa cuando la prevalencia real es baja (aproximadamente menor a 0,2) (Figura 5).

Figura 5. Sensibilidad a nivel de grupo en función de la prevalencia real de la enfermedad y el número de MC a considerar (número de muestras/ MC= 5, sensibilidad=0,647, especificidad=0,981).



Dentro de esta estrategia se puede identificar un caso particular, que es aquel en el cual se toma una sola MC a partir de todos los animales que conforman el grupo a evaluar. Las pruebas del anillo en leche y ELISA para diagnóstico de brucelosis o la toma de una muestra de leche del tanque de enfriamiento para recuento de diferentes

grupos microbianos con la finalidad de diagnosticar la calidad sanitaria de esa leche, son algunos de los ejemplos más comunes en la práctica veterinaria. En este caso sólo se puede concluir si el grupo es positivo o negativo y su interpretación es similar a la que se realiza cuando se toman muestras individuales para cada animal.

Analicemos el siguiente ejemplo sobre el uso de una sola MC para el diagnóstico de enfermedades a nivel grupal. Los programas de control y erradicación de tuberculosis se basan en la eliminación de todos los animales que resultan positivos a la prueba intradérmica de la PPD. Aquellos establecimientos que hayan logrado la certificación oficial de «establecimientos libres de tuberculosis», deben ser auditados mediante la aplicación de la PPD a 15 animales y evaluar la reacción a las 72 horas posteriores. Esta estrategia es costosa y demanda mucho tiempo para ser aplicada, por lo que se ha evaluado el uso de técnicas moleculares (PCR) para el diagnóstico de tuberculosis bovina en leche de tanque como alternativa para simplificar las auditorías (Soutullo et al., 2006). La concordancia entre la técnica molecular aplicada a una sola MC de leche y la prueba tradicional de intradermorreacción es baja (Índice Kappa= 0,106). Diferentes factores pueden identificarse como responsables de esta baja concordancia:

- a) no todos los animales reactantes positivos a la PPD son eliminadores de bacilos por leche;
- b) aún los animales que eliminan bacilos por leche lo hacen de manera cíclica;
- c) es posible que los animales reactantes positivos a la PPD no se encuentren en lactancia al momento de realizar la prueba;
- d) efecto de dilución de la leche en el tanque de enfriamiento, especialmente si la prevalencia de la enfermedad es baja;
- e) no obstante, esta técnica puede ser empleada como parte de un programa de vigilancia de establecimientos considerados libres de la enfermedad ya que se asume que la especificidad de la PCR es cercana al 100 %.

3. Muestreo de poblaciones específicas

Otra alternativa disponible para incrementar la sensibilidad a nivel de grupo, especialmente cuando la prevalencia real de la enfermedad es baja, es el muestreo dirigido hacia poblaciones que, *a priori*, tienen mayor probabilidad de presentar la enfermedad. Este tipo de estrategias es empleada cuando el objetivo es detectar la presencia de la enfermedad en un grupo y cuando se han identificado ciertos factores de riesgo (edad, sexo, raza) que permitirían agrupar a los animales en diferentes estratos. La información adicional disponible sobre estos grupos de animales que comparten alguna característica que los hace más susceptibles a la enfermedad colabora para reducir el número de muestras a tomar.

Uno de los ejemplos típicos de este tipo de muestreos es el que se aplica para la detección de la Encefalopatía Espongiforme Bovina (EEB) o «mal de las vacas locas».

El monitoreo de la enfermedad se realiza a nivel de frigoríficos. Los factores que dirigen el muestreo son: animales con más de 36 meses de edad (dado que es una enfermedad de curso crónico y no se han detectado lesiones encefálicas en bovinos menores de esa edad), animales con signología nerviosa o que hayan sido sometidos a sacrificio de urgencia. De esta forma se reduce el número de muestras a tomar (una gran ventaja cuando los métodos diagnósticos son onerosos) y se centra la atención en aquellos grupos de animales con mayor probabilidad de presentar la enfermedad.

4. Programas de computación

Los cálculos relacionados con la sensibilidad y especificidad a nivel de grupo pueden ser complejos de realizar. Para ello se han desarrollado diversos programas de computación que son de gran ayuda.

El programa HERDACC está disponible en la página <http://epiweb.massey.ac.nz> y se emplea para obtener valores de sensibilidad y especificidad a nivel de grupo para diferentes combinaciones de prevalencias, sensibilidad y especificidad de la prueba diagnóstica, número de animales a muestrear y valores umbrales.

FREECALC es una calculadora epidemiológica para el diseño y análisis de investigaciones desarrolladas para detectar enfermedad o para definir la ausencia de la misma en una población. El programa puede obtenerse a partir de la misma página que el programa anterior o desde <http://www.pnc.com.au/~angus>

Para aquellos lectores que deseen profundizar en la interpretación de las pruebas diagnósticas a nivel de grupo, se recomienda la lectura de Christensen y Gardner (2000).

Bibliografía

- Houe, H., Kjaer Ersboll, A., Toft, N., Agger, J.F. (2003).** Veterinary epidemiology: from hypothesis to conclusion. Ed. The Royal Veterinary and Agricultural University, Frederiksberg, Dinamarca, 340 pp.
- Christensen, J. y Gardner, I.A. (2000).** Herd-level interpretation of test results for epidemiologic studies of animal diseases. *Preventive Veterinary Medicine*, **45**: 83–106.
- Dohoo, I., Martin, W., Stryhn, H. (2003).** *Veterinary Epidemiology Research*. Ed. AVC Inc., Charlottetown, Prince Edward island (Canadá), 706 pp.
- Gardner, I.A., Greiner, M. (1999).** Advanced Methods for test Validation and Interpretation in Veterinary Medicine Freie. Univ. Berlin, 74 pp.
- Soutullo, A., Zumárraga, M., Capello, G., García, M.I., Marini, R., López, M., Abdala, A., Tarabla, H., Zerbini, E., Latini, M., Cataldi, A., Canal, A. (2006).** Utilidad de la PCR en el diagnóstico de tuberculosis en leche de tanque de tambos de la cuenca lechera santafecina. En XVI Reunión Científico Técnica de la AAVDL. Mar del Plata (Argentina), 5–7 de diciembre de 2006.

Capítulo 13

Aproximación probabilística a problemas con las pruebas diagnósticas

Hasta este momento el libro trató de ofrecer conceptos claros y simples con el propósito de facilitar la interpretación de los resultados obtenidos luego de la aplicación de pruebas diagnósticas, sin profundizar en los aspectos matemáticos o probabilísticos que sustentan a esta rama de la epidemiología.

El veterinario suele ver las enfermedades como estrictas relaciones causa/efecto. «Si el resultado de una prueba diagnóstica es positivo entonces el animal presenta la enfermedad». En los diferentes capítulos se ha desarrollado una serie de conceptos que deberá ayudar al profesional a comprender que el axioma arriba expuesto no siempre es real y que no puede rehusarse a aceptar que las pruebas diagnósticas siempre ofrecerán resultados falsos positivos y/o falsos negativos.

Un aspecto adicional que no fue abordado completamente se refiere a visualizar la aplicación de las pruebas diagnósticas como fenómenos probabilísticos. Con la intención de simplificar este punto y no incursionar en demasía en los fundamentos estadísticos, se desarrollará a través de un ejemplo.

1. Desarrollo de un ejemplo

Supongamos que Argentina desea importar 1000 equinos. El país de procedencia de los animales presentó algunos casos de una enfermedad que nuestro país está tratando de erradicar en los últimos 5 años y se estima que la prevalencia de la enfermedad es, en promedio, del 2,5 %. El servicio sanitario animal decide entonces, antes de que los animales ingresen al país, aplicarles una prueba diagnóstica que fue seleccionada por poseer una alta sensibilidad ya que desea minimizar los falsos negativos. La prueba tiene una sensibilidad del 99 % y una especificidad del 95 %.

Sin embargo, el servicio sanitario sigue dudando sobre la importación en estas condiciones y se pregunta: ¿cuál es la probabilidad de que un animal que porte la enfermedad ingrese al país sin ser detectado por la prueba diagnóstica (falso negativo)? Y además, como máximo, ¿cuántos equinos falsos negativos podrían ingresar?

Veamos a continuación dos maneras de resolver el problema.

1.1. Aproximación determinística

La aproximación determinística o puntual es la que estuvimos trabajando a lo largo de este libro y que básicamente sigue el siguiente razonamiento:

Se sabe que ingresarán a nuestro territorio 1000 equinos y que la prevalencia (PR) promedio de la enfermedad en el país de origen es de 2,5 %, por lo que el número de animales que portando la misma se calcula de este modo:

Animales enfermos = Animales a importar x Prevalencia estimada de la enfermedad

$$\text{Animales enfermos} = 1000 \times 0,025 = \mathbf{25 \text{ animales con la enfermedad}}$$

Ahora deseamos saber cuántos animales falsos negativos podrán ingresar al país aplicando una prueba diagnóstica con 99 % de sensibilidad y 95 % de especificidad. El número de falsos negativos se calcula entonces de la misma manera que lo estuvimos haciendo hasta el momento:

n° falsos negativos = n° Animales enfermos x (1 – sensibilidad de la prueba diagnóstica)

$$n^{\circ} \text{ falsos negativos} = 25 \times (1 - 0,99) = \mathbf{0,25 \text{ animales}}$$

Es decir, con esta aproximación la probabilidad de que un animal dé un resultado falso negativo es sumamente baja, por lo que, aparentemente, no ingresaría ningún animal falso negativo:

Probabilidad de falso negativo = Prevalencia de la enfermedad x (1 – sensibilidad)

$$\text{Probabilidad de falso negativo} = 0,025 \times (1 - 0,99) = \mathbf{0,00025}$$

El proceso anterior puede ser traducido a una tabla de 2 x 2 a las que ya estamos acostumbrados (Cuadro 1):

Cuadro 1. Clasificación de una población de equinos a importar con respecto a su estado de salud (prevalencia real del 2,5 %) luego de aplicar una prueba diagnóstica (Sens.= 99 %, Esp.= 95 %)

Condición	Enfermos	Sanos	Total
Test positivos	24,75	48,75	73,50
Test negativos	0,25	926,25	926,50
Total	25,00	975,00	1000,00

Ahora bien, si empleamos los conceptos que introducimos en el Capítulo 12 sobre la interpretación de pruebas diagnósticas, aplicadas a nivel de rodeo, podemos decir que la probabilidad de que un animal dé un resultado falso negativo también puede ser expresada así :

$$P(E+ | T-) = \frac{PR \times (1 - \text{Sens.})}{PR \times (1 - \text{Sens.}) + (1 - PR) \times \text{Esp.}}$$

$$P(E+ | T-) = 0,000269$$

y de esta forma podemos calcular la probabilidad de que al menos un animal presente la enfermedad, puesto que todos los animales analizados (n= 1000) dieron un resultado negativo (todosT-) cuando se les aplicó la prueba diagnóstica (Vose, 1997):

$$P(E+ \geq 1 | \text{todosT-}) = 1 - [1 - P(E+ | T-)]^n$$

$$P(E+ \geq 1 | \text{todosT-}) = 1 - [1 - 0,000269]^{1000}$$

$$P(E+ \geq 1 | \text{todosT-}) = 0,2365$$

Es decir, la probabilidad de que un animal dé un resultado falsamente negativo es de aproximadamente 0,00025; y la probabilidad de que en el lote de 1000 equinos a importar haya al menos un animal falso negativo es del 23,65 %.

Independientemente del análisis que se haga de la información (individual o a nivel de rodeo), el resultado final de este proceso genera estimados de tipo puntual, ya que tanto las variables de entrada (prevalencia de la enfermedad y sensibilidad de la prueba diagnóstica) como de salida (animales falsos negativos) están expresadas como números simples o valores puntuales (Signorini, 2008).

1.2. Aproximación estocástica

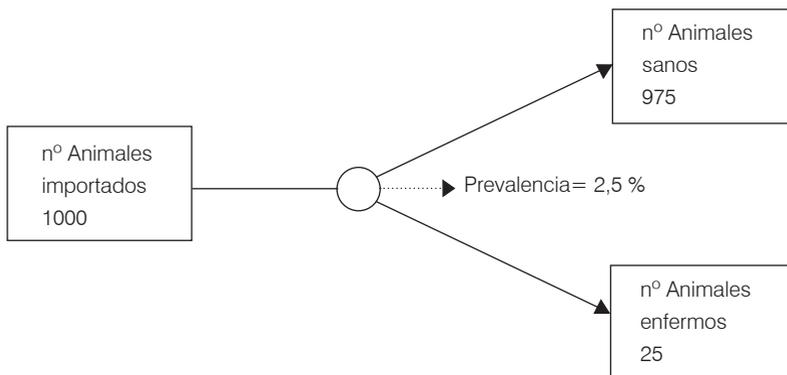
Este método parte de la premisa de que la prevalencia no sólo puede concebirse como la proporción de animales enfermos dentro de una población específica sino

que se puede interpretar como la probabilidad de seleccionar un animal enfermo en una población objetivo (Signorini, 2008; Vose, 1997).

Antes de continuar con el desarrollo del ejemplo, vamos a analizar otro ejemplo simple para que el concepto quede claro. Si acude a su veterinaria un cliente que tiene una perra preñada y le pregunta cuál es la probabilidad de que un cachorro sea macho, usted rápidamente responde que es del 50 %. Acorde a esto, el cliente le pregunta cuántos cachorros puede tener esa perra (según la raza y edad del animal) y cuántos espera que sean macho. Luego de practicarle una ecografía, usted responde que la perra tiene 10 cachorros en el útero y que por lo tanto espera que 5 de ellos sean machos. Para calcular lo anterior usted multiplicó el número de cachorros (10) por la probabilidad de que estos sean machos (50 % o 0,5) y obtuvo el valor final (5 cachorros machos). Con esta información, el cliente le pregunta, ¿es probable que los 10 cachorros sean machos? Para responder, usted tendrá que hacer un poco de esfuerzo para recordar sus clases de estadística y, si bien puede que no responda exactamente a la probabilidad que ese evento posee (0,097 %), seguramente contestará que, si bien es probable que la totalidad de los cachorros sea macho, es poco probable que eso ocurra. De hecho, es posible que todos los cachorros sean machos o que todos sean hembras y todas las posibilidades intermedias, la única diferencia es la probabilidad de ocurrencia de cada una de esas combinaciones. Intuitivamente, pensamos que 5 de esos cachorros van a ser machos porque hallamos el valor esperado o más probable de ese evento, pero no analizamos todos los resultados posibles.

Algo similar ocurre en el caso que estamos analizando; sabemos que ingresarán 1000 animales y que la prevalencia en el país de origen es del 2,5 % y rápidamente hallamos el valor de 25 animales enfermos que ingresarán potencialmente al país, pero sólo obtuvimos el valor esperado y no consideramos todos los posibles resultados (Figura 1).

Figura 1. Gráfico de árbol para el cálculo puntual del número de animales enfermos dentro de la población de equinos a importar

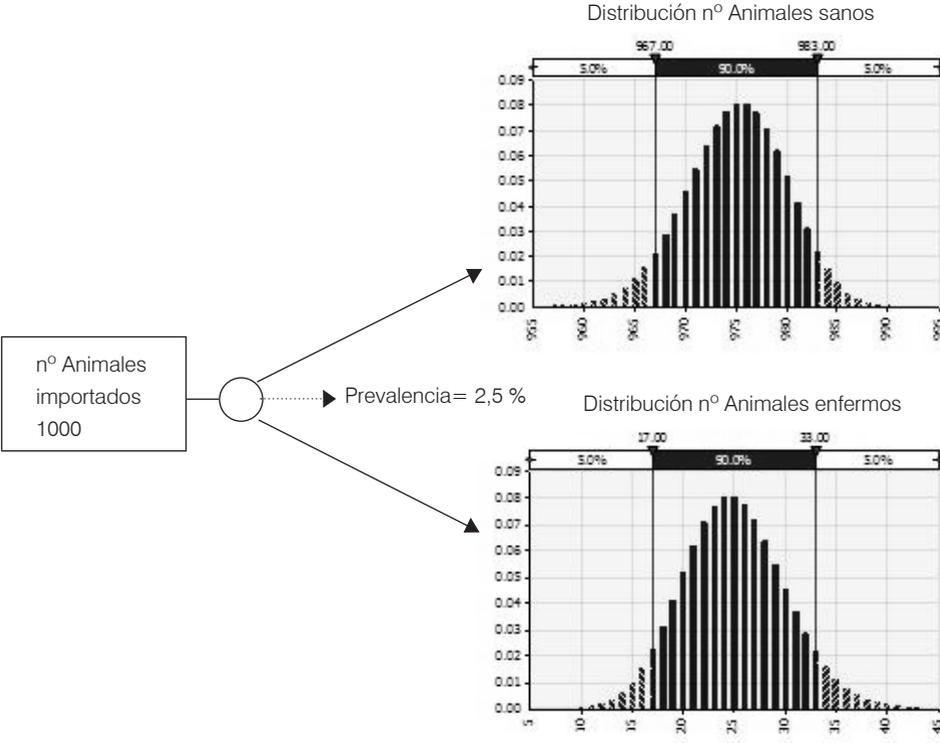


El desafío entonces es analizar el problema desde un punto de vista estadístico, de manera tal que podamos contemplar la totalidad de los eventos posibles y no solamente el más probable o esperado. Teniendo en cuenta que el evento que estamos considerando tiene dos resultados factibles mutuamente excluyentes (el equino presenta o no presenta la enfermedad) y que la condición de un animal es independiente de la condición sanitaria de cualquiera de los otros animales que conforman la población objetivo, el número de animales enfermos dentro de esta población sigue una distribución binomial, con parámetros n (número de animales en la población) y p (prevalencia de la enfermedad en la población), siendo en nuestro estudio:

$$n^{\circ} \text{ de animales enfermos} \sim \text{Binomial} (1000; 0,025)$$

Es decir, el número de animales enfermos sigue una distribución probabilística de tipo binomial. A diferencia del método determinístico, que sólo da un resultado posible (valor puntual), el método estocástico ofrece una serie de resultados que en conjunto siguen una distribución binomial (Figura 2), incorpora así la variabilidad natural de la enfermedad a la población considerada. (Signorini y Tarabla, 2009; Olvera-Yabur, et al., 2010)

Figura 2. Gráfico de árbol para el cálculo probabilístico del número de animales enfermos dentro de la población de equinos a importar.



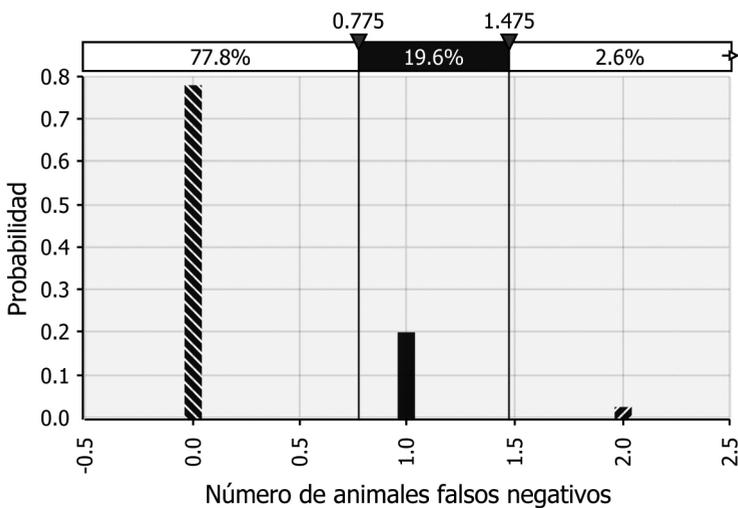
Las medias de las distribuciones del número de animales sanos (975) y enfermos (25) serán los mismos valores que los obtenidos a través de la metodología determinística, ya que el valor medio de una distribución binomial es igual al número de animales por la probabilidad de presentar la enfermedad ($n \times p$). Pero, como se observa en la Figura 2., la aproximación estocástica ofrece todos los posibles valores posibles para cada categoría asociada a la probabilidad de ocurrencia. Por ejemplo, se estima que el número esperado de animales enfermos será de 25, pero el intervalo de confianza del 90 % nos indica que puede haber entre 17 y 33 animales enfermos.

Ahora, conocemos la distribución de animales que son portadores de la enfermedad y queremos calcular el número de animales que, estando enfermos, no son diagnosticados como positivos por la prueba diagnóstica. Al igual que en la aproximación determinística, se emplea la sensibilidad de la prueba diagnóstica (99 %), pero aquí lo haremos como una probabilidad, ya que cumple con los supuestos de la probabilidad binomial (genera dos resultados posibles, mutuamente excluyentes e independientes). Entonces, se calcula el número de animales falsos negativos como una función de distribución binomial, con parámetros n (distribución del número de animales enfermos) y p (sensibilidad de la prueba diagnóstica):

$$\text{Falsos negativos} \sim \text{Binomial} [\text{distribución de animales enfermos}, (1 - 0,99)]$$

De esta forma, el número de animales falsos negativos no es un valor único en este caso sino que es una distribución de probabilidad que considera todos los eventos posibles y su probabilidad asociada (Figura 3).

Figura 3. Distribución de la probabilidad del número de equinos falsos negativos



De acuerdo con la Figura 3, es altamente probable que no ingrese equino falso negativo alguno al país, dado que en el 77,8 % de las veces que se analicen animales enfermos en su totalidad serán detectados como positivos. No obstante, el ingreso de animales falsos negativos puede darse porque, en promedio, en el 19,6 % de los casos puede ocurrir que 1 animal no sea detectado y que en el 2,6 % de las veces sean 2 los animales falsos negativos que superan la prueba. Es decir, con esta metodología podemos estimar que la probabilidad de que ingrese al menos un equino falso negativo al país es del 22,2 %, pero esta probabilidad puede variar (con una confianza del 90 %) entre 15,7 % y 28,2 %. Seguramente usted ya habrá notado que la probabilidad estimada de encontrar al menos un equino falso negativo en el rodeo es muy similar para las dos aproximaciones: determinística (23,65 %) y estocástica (22,2 %).

Luego de aplicar este enfoque, la respuesta que le daremos al servicio de sanidad animal en términos cuantitativos no será muy diferente de la que aportemos si usamos la aproximación determinística (Signorini et al., 2009). No obstante, la gran ventaja que nos ofrece analizar estos eventos de manera probabilística radica en que nos permite contemplar todos los escenarios posibles y ponderarlos conforme a la probabilidad de ocurrencia. Posiblemente, con esta información el servicio sanitario preste mayor atención a la solicitud de importación de equinos de este país porque el ingreso de animales portadores de la enfermedad no puede descartarse y de hecho es un evento factible cuya probabilidad real se incrementará en la medida en que aumente el número de animales a importar (Signorini y Tarabla, 2010).

Este tipo de enfoques no se realiza rutinariamente; de hecho, son casos excepcionales en los cuales se recurre a ellos. Pero cuando se trata de evaluar procesos muy complejos, con desenlaces que pueden generar un grave impacto en la salud pública o en la sanidad animal, o cuando se desean tomar acciones de control, se emplea esta herramienta que en conjunto se llama "Evaluación cuantitativa de riesgos". Aunque poco utilizado en el área de pruebas diagnósticas, este enfoque es ampliamente empleado en la evaluación de procesos (Signorini, 2008; Signorini y Tarabla, 2009; Signorini et al., 2009; Signorini y Tarabla, 2010; Signorini et al., 2012; Olvera-Yabur et al., 2010).

En nuestro ejemplo, al asumir la prevalencia de la enfermedad y la sensibilidad de la prueba diagnóstica como variables estocásticas, estamos reconociendo la variabilidad natural de la enfermedad y la variabilidad en el desempeño de las pruebas diagnósticas utilizadas. Si bien este capítulo no pretende profundizar en la incorporación de la incertidumbre a las variables de entrada de un modelo, podemos afirmar que es factible reemplazar los valores determinísticos por distribuciones de probabilidad que simulen la variabilidad real del proceso. Por ejemplo, podemos preguntarnos si el número de equinos a importar es siempre y exactamente 1000 animales o si éste no es un valor promedio. También dijimos que la prevalencia de la enfermedad en el país de origen era, en promedio, de 2,5 %, pero no estamos totalmente seguros de que ése

sea el valor real en la región específica de proveniencia de los animales. De la misma manera podemos considerar que la sensibilidad de la prueba diagnóstica puede variar según la forma en la cual se la realice o entre operadores, lo que nos obligará a incorporar esta variabilidad en el parámetro. Como se advierte, los modelos se pueden complejizar tanto como uno desee y, fundamentalmente, en función de la información que uno disponga.

Bibliografía

- Signorini, M.L. (2008).** Uso de dos metodologías en la estimación del riesgo de animales falsos negativos y prevalencia aparente de *Escherichia coli* O157:H7. Revista FAVE–Ciencias Veterinarias, 7(1–2): 71–82.
- Signorini, M. & Tarabla, H. (2009).** Quantitative risk assessment for Verocytotoxigenic *Escherichia coli* in ground beef hamburgers in Argentina. Int. J. Food Microbiol., 132: 153–161.
- Signorini, M., Marín, V.; Quinteros, C. & Tarabla, H. (2009).** Hábitos de consumo de hamburguesas y riesgo de exposición a *Escherichia coli* VTEC. Rev. Argentina de Microbiología, 41(3): 168–176.
- Signorini, M. & Tarabla, H. (2010).** Interventions to reduce verocytotoxigenic *Escherichia coli* in ground beef in Argentina: A simulation study. Preventive Veterinary Medicine, 94: 36–42.
- Signorini, M.L., Gaggiotti, M., Molineri, A., Chiericatti, C.A., Zapata de Basílico, M.L., Basílico, J.C., Pisani, M. (2011).** Exposure assessment of mycotoxins in cow's milk in Argentina. Food and Chemical Toxicology, 50: 250–252.
- Olvera-Yabur, A.; Signorini, M.L. & Tarabla, H. (2010).** *Escherichia coli* verotoxigénica: Modelo cuantitativo de exposición y escenarios de riesgos en canales bovinas de Argentina. Revista Panamericana de Salud Pública, 27(6): 403–413.
- Vose, D. (1997).** Risk analysis in relation to the importation and exportation of animal products. Rev. Sci. Tech. Off. Int. Epiz., 16(1): 17–29.

Capítulo 14

Elección de la prueba

¿Que criterio se debe utilizar para seleccionar la prueba diagnóstica más apropiada? Antes que nada, hay que recordar lo siguiente:

1. Todas las pruebas diagnósticas son imperfectas (no hay ninguna prueba que sea, al mismo tiempo, 100 % sensible y 100 % específica).
2. Las evaluaciones de sensibilidad y especificidad realizadas en el laboratorio o condiciones experimentales usualmente sobrestiman el real desempeño de la prueba a campo.
3. Las modificaciones técnicas que se realicen para mejorar la sensibilidad de la prueba pueden resultar en una disminución en la especificidad y viceversa.
4. Un punto de corte no es necesariamente óptimo para todas las situaciones. La elección de un valor adecuado depende, a su vez, de varios factores que incluyen el objetivo del diagnóstico, los objetivos del usuario, el costo de los errores de clasificación de los individuos en enfermos y no enfermos y si los resultados se interpretarán a nivel individual o grupal (Tarabla, 1995).

En la detección de antibióticos en leche, por ejemplo, la industria láctea está interesada en métodos de análisis de bajo costo, fácil realización y rápidos resultados. El objetivo diagnóstico es detectar todos los rodeos positivos. La prueba debe ser altamente sensible dado que el mayor costo va a provenir de los falsos negativos que puedan filtrarse y contaminar un volumen mayor de producto. Por el contrario, para el productor lechero el mayor interés está en una prueba confiable y precisa. Ésta deberá ser altamente específica puesto que el mayor costo estará representado por los diagnósticos falsos positivos que pueden resultar en penalidades inmerecidas para su explotación agropecuaria. Esta dicotomía industria/productor incluso influenciará la elección de una prueba confirmatoria. En el primer caso bastará con un método micro-

biológico que detecte cualquier inhibidor en la leche, sea o no una sustancia prohibida. No obstante, para el segundo será más adecuada la cromatografía líquida de alto rendimiento que puede detectar específicamente la presencia antibiótica.

¿Qué se debe priorizar, una alta sensibilidad, una alta especificidad, el valor predictivo más elevado o la mayor eficiencia? Como se describió a lo largo de este texto, no es posible tenerlo todo simultáneamente. El valor predictivo describe cómo puede funcionar una prueba en una población pero da muy poca información sobre su poder discriminatorio entre sanos y enfermos. La selección de una prueba se debe hacer en función de su poder discriminatorio y no de su valor predictivo (Robertson, 1963), y es preciso seleccionar la prueba disponible que acredite la mayor sensibilidad y especificidad. En general, en los diagnósticos en casos individuales es muy importante que la prueba sea muy sensible para detectar todo individuo enfermo. En un resultado falso positivo se pueden utilizar pruebas complementarias para descartar esa posibilidad (Marchevsky, 1974).

Un alto valor predictivo sólo es factible que sea utilizado como ayuda en la selección de la prueba en algunos casos en los cuales el tratamiento de los falsos positivos puede tener consecuencias serias para la salud del paciente (ej.: en medicina humana, radiaciones en pacientes sanos a los cuales se les diagnostique cáncer) (Galen y Gambino, 1975:237). Asimismo, una prueba podría ser seleccionada por su alta eficiencia en casos de enfermedades serias pero tratables en las que es posible que causen igual daño los diagnósticos falsos positivos como los falsos negativos (ej.: infarto de miocardio en Medicina Humana).

Por otra parte, cuando se utilicen pruebas diagnósticas individuales para diagnosticar enfermedad a nivel grupal (ej.: rodeos), se deben seleccionar pruebas altamente específicas. Las variaciones moderadas en su sensibilidad no tendrán gran impacto en la sensibilidad a nivel grupal (Duffy, 1994:82).

En el caso de estudios poblacionales donde el interés esté dado en la estimación de una prevalencia y no en una toma de decisión a nivel individual, los falsos positivos y los falsos negativos no tienen una consecuencia directa sobre los individuos que componen la población pero sí en la sobre o subestimación del parámetro poblacional. Aunque una moderada falta de sensibilidad no es de gran importancia, sí lo es la falta de especificidad (ej.: la proporción de falsos positivos), porque puede llevar a una sobrestimación de la prevalencia real. Esta prevalencia inflada quizá sobredimensione la magnitud del problema, especialmente en los casos en que se toman medidas sanitarias de alto costo como el sacrificio de animales infectados. De cualquier manera, la selección del método diagnóstico a utilizarse para estudios poblacionales no sólo debe basarse en la sensibilidad y especificidad de la prueba sino también en su simpleza y su costo (Marchevsky, 1974).

En el caso de la aplicación de pruebas en programas de control y erradicación de enfermedades para prevenir el ingreso de animales enfermos en regiones o grupos de rodeos libres son especialmente necesarios pruebas de alta sensibilidad para que no

pase ningún falso negativo. La especificidad aquí tiene un rol secundario (Astudillo y Kantor, 1981). En el comienzo del programa se trata de reducir los falsos negativos a expensas de un aumento en los falsos positivos para evitar que queden animales infectados no detectados que diseminen la enfermedad. No obstante, cuando el programa se acerca a su culminación y se ha logrado una baja prevalencia de la enfermedad, se puede requerir un cambio en la estrategia (Martin, 1984). Este cambio implica generalmente la utilización de una nueva prueba, aunque en algunos casos pueda lograrse mediante la elección de un nuevo punto de corte en la misma prueba diagnóstica para darle una mayor especificidad (Barajas-Rojas et al., 1993). Por otra parte, para detectar la presencia de enfermedad en áreas libres y bien protegidas son necesarios pruebas altamente específicas para minimizar casos de reactores aislados falsos positivos (Astudillo y Kantor, 1981). Realizar una prueba tamiz con alta sensibilidad en una población con baja prevalencia de la enfermedad puede ser una pérdida de dinero y tiempo. Las pruebas tamices deben realizarse en poblaciones con alta prevalencia o a grupos de riesgo (Gordis, 2004:333).

De cualquier modo, la aceptabilidad de cualquier opción de prueba diagnóstica a utilizar dependerá sobre todo del costo de dejar un animal infectado en un rodeo comparado con el costo de sacrificar animales falsos positivos. Esta disyuntiva entre el costo de los falsos negativos *versus* el costo de los falsos positivos se reiterará en todas las evaluaciones que se realicen para la elección de la prueba diagnóstica más adecuada para cada situación.

Bibliografía

- Astudillo, V.M., Kantor, I.N. (1981).** El problema de la validez de una prueba diagnóstica para uso masivo como procedimiento estadístico de clasificación. *Bltn. Centr. Panam. Fiebre Aftosa.* **43-44:** 37-43.
- Barajas-Rojas, J.A., Riemann, H.P., Franti, C.E. (1993).** Notes about determining the cut-off value in enzyme linked immunosorbent assay (ELISA). *Prev. Vet. Med.* **15:** 231-233.
- Duffy, S. (1994).** Diagnosis at herd level. *Proc. VIIth Internat. Symposium Vet. Lab. Diag., Soc. Med. Vet. and Asoc. Arg. Vet. Lab. Diag., Buenos Aires,* p. 82.
- Galen, R.S., Gambino, S.R. (1975).** *Beyond Normality, the Predictive Value and Efficiency of Medical Diagnoses.* John Wiley & Sons, New York, 237 pp.
- Gordis, L. (2004).** *Epidemiology.* 3º Edición. Ed. Elsevier Saunders, Philadelphia, Pennsylvania (Estados Unidos de América), 333 pp. ISBN 1-4160-2530-8.
- Marchevsky, N. (1974).** Errores en las estimaciones de prevalencia en estudios de población: Un método para calcular la prevalencia real. *Zoonosis.* **16:** 81-109.
- Martin, S.W. (1984).** Estimating disease prevalence and the interpretation of screening test results. *Prev. Vet. Med.,* **2:** 463-472.
- Robertson, T.G. (1963).** Diagnosis of bovine tuberculosis. The evaluation of tuberculin tests. *Vet. Rec.* **105:** 226-228.
- Tarabla, H.D. (1995).** Criterios para la elección de una prueba diagnóstica. Resúmenes 11 Congr. Arg. y 11 Congr. Latinoam. *Zoonosis,* Buenos Aires, Argentina, p. 208.