

CAPÍTULO 1.1.6.

VALIDACIÓN DE LAS PRUEBAS DE DIAGNÓSTICO DE LAS ENFERMEDADES INFECCIOSAS DE ANIMALES TERRESTRES

INTRODUCCIÓN

Una validación y verificación adecuadas de las características de rendimiento de las pruebas de diagnóstico de las enfermedades infecciosas son fundamentales para garantizar que tales pruebas se apliquen e interpreten de forma científicamente sólida y defendible (Colling y Gardner, 2021). Desde que se adoptó por primera vez en 1996, el Procedimiento de Desarrollo y Validación de Pruebas de la OMSA (Figura 1) ha actuado como norma reconocida internacionalmente para la validación de pruebas de diagnóstico veterinario de enfermedades infecciosas.

La validación es un proceso que determina la idoneidad de una prueba¹, que se ha desarrollado, optimizado y estandarizado adecuadamente para una finalidad definida. Todas las pruebas de diagnóstico (con independencia de si se utilizan en el laboratorio o a pie de establo) deben validarse para la especie y el tipo de muestra en los que se utilizarán. La validación incluye estimaciones de las características de rendimiento analítico y diagnóstico de una prueba. En el contexto de este capítulo, una prueba que ha superado las tres primeras fases de la validación (véase la Figura 1, abajo), incluida la caracterización del rendimiento, puede designarse como “validada para la finalidad inicial definida(s)”. Pero para que una prueba siga estando validada, su rendimiento debe monitorizarse en detalle en las condiciones habituales de uso, a menudo controlando el comportamiento de los controles de la prueba dentro de cada ejecución, y realizando una evaluación exhaustiva durante el proceso durante su uso sistemático para el diagnóstico en la población de destino y a lo largo del tiempo. En el caso de que empiece a dar resultados que no concuerden con los datos de la validación original, la prueba podría no ajustarse a la/s finalidad/es definida/s.

Las pruebas realizadas en individuos o en poblaciones tienen varias finalidades, tales como: documentar la ausencia de una determinada enfermedad en un país o región, evitar su propagación a través del comercio, contribuir a erradicar una infección de una zona o país, confirmar el diagnóstico de los casos clínicos, estimar la prevalencia de una infección para facilitar el análisis del riesgo, identificar a los animales infectados con vistas a implementar medidas de control, y clasificar los animales según su salud o estado inmunitario tras la vacunación. Una única prueba puede validarse para una o varias finalidades definidas optimizando sus características de rendimiento para cada una, como, por ejemplo, fijando una alta sensibilidad diagnóstica (Dse), asociada a una baja especificidad diagnóstica (Dsp) para una prueba de cribado, o, por el contrario, fijando una DSP alta asociada a una Dse más baja para una prueba confirmativa (véase la Sección A.1 Definición de la/las finalidad/es de una prueba).

Este capítulo se centra en los criterios que deben cumplirse durante la realización y la validación de pruebas de cualquier tipo y en la métrica que se utiliza para definir el rendimiento de una prueba. La inclusión de la ejecución de la prueba como parte del proceso de validación de esta puede parecer contra-intuitiva, pero en realidad, tres de los criterios de validación requeridos (definición de la/s finalidad/es definida/s, optimización y estandarización) que deben evaluarse para llegar a la validación de una prueba incluyen pasos de su proceso de ejecución. Según esto, el proceso de ejecución de la prueba da lugar a un sistema de validación de esta, y ambos contienen criterios de validación que deben cumplirse. Los principios orientativos que se describen aquí también son aplicables a las enfermedades infecciosas no incluidas en la lista de la OMSA. Se aportan otras

1 “Prueba,” “método analítico,” y “test” son sinónimos a los efectos de este capítulo, y por lo tanto se utilizan indistintamente.

directrices en una serie de Recomendaciones de la OMSA para la validación de pruebas de diagnóstico (capítulos 2.2.1 a 2.2.8 del Manual Terrestre²) que se adaptan a varios tipos de prueba que en esencia son distintos (por ejemplo, la detección de anticuerpos, antígenos o ácidos nucleicos) y aportan más información sobre temas específicos relacionados con la validación de pruebas de diagnóstico. Para información específica relativa a especies de fauna salvaje, consúltese el capítulo 2.2.7 Principios y métodos para la validación de las pruebas de diagnóstico de las enfermedades infecciosas aplicables a la fauna salvaje (Jia et al., 2020; Michel et al., 2021). La información que se aporta en el capítulo, que es específica de especies de fauna salvaje, también podría ser de utilidad para la validación de pruebas realizadas en animales domésticos, por ejemplo, cuando se disponga de un escaso número de muestras.

En el número de la Revista Científica y Técnica de la OMSA titulado Ciencia de Validación de las Pruebas de Diagnóstico (Vol. 40, abril de 2021) se ofrece una recopilación actualizada de las normas de validación pertinentes (pertenecientes o no a la OMSA) y documentos de orientación sobre todas las fases de la validación y de los estudios de competencia de las pruebas de diagnóstico, incluidos el diseño, el análisis y la interpretación, así como estudios sobre comunicación clara y de casos. Existen normas publicadas para la presentación de informes revisados por pares de estudios sobre la exactitud del diagnóstico (STARD) de enfermedades infecciosas humanas o de animales terrestres (paratuberculosis, modelo bayesiano de clases latentes) (Bossuyt et al., 2015; Gardner et al., 2011; Kostoulas et al., 2017) o acuáticos (Gardner et al., 2019; Kostoulas et al., 2021). Se describen brevemente estudios de verificación (Kirkland y Newberry, 2021) y de comparabilidad (Reising et al., 2021) al final de este capítulo para los ensayos que han completado al menos la fase 2 de la vía de la OMSA. Existe una necesidad apremiante de desarrollar directrices y normas de validación para el uso cada vez mayor de pruebas a pie de establo (POCT, por las siglas en inglés de point-of-care tests). Normalmente, los POCT se utilizan sobre el terreno en condiciones ambientales variables, en una gama de tipos de muestras recogidas en entornos no estériles, por parte de personal con diversidad de grados de experiencia, formación y competencia. Las condiciones en las que se obtienen las pruebas sobre el terreno, incluidas las variaciones extremas de temperatura y humedad, así como otras variables, como la calidad del agua y de los reactivos, una cadena de frío inadecuada, la habilidad del personal y la posible deficiencia o ausencia de sistemas de garantía de calidad pueden contribuir a que la exactitud de los resultados de las pruebas sea inferior a la indicada por los fabricantes de las POCT o a la obtenida en un laboratorio acreditado (véase la Figura 1, Fase 5). Las consecuencias de un resultado positivo y la necesidad de que un laboratorio acreditado realice pruebas de confirmación y elabore informes, en particular cuando se realiza una prueba para una enfermedad exótica, deben integrarse en las políticas y directrices existentes en materia de pruebas. Las normas y recomendaciones específicas sobre las POCT, como el listado de verificación de evidencias clave a pie de establo (POCKET) para la notificación de evidencias multidimensionales; las tarjetas de puntuación y las directrices para la evaluación de pruebas POCT; las directrices sobre prácticas de calidad en pruebas POCT no instrumentadas (es decir, los que no requieren un equipo específico); y la norma ISO/TS 22583:2019 Orientación para supervisores y operadores de dispositivos para pruebas POCT proporcionan orientación para el personal sanitario. En algunos países, como Alemania, las POCT que detectan enfermedades animales de declaración obligatoria requieren una autorización formal por parte de la autoridad nacional de concesión de licencias. En otros países, se fomenta la acreditación de las POCT ante organizaciones como la Organización Mundial de Sanidad Animal (fundada como la OIE) y las autoridades nacionales para las pruebas de diagnóstico, pero no es obligatoria (Halpin et al., 2021; Hobbs et al., 2020; y sección 2.5 Robustez y resistencia, más adelante).

CONSIDERACIONES PRELIMINARES EN EL DESARROLLO Y VALIDACIÓN DE PRUEBAS

Todos los laboratorios deben cumplir los requisitos establecidos en el Capítulo 1.1.5 Gestión de la calidad en los laboratorios de pruebas veterinarias (Newberry y Colling, 2021). Con ello, se minimizará la influencia de factores que no dependen de la prueba en sí, como el instrumental, errores del operario, la elección del reactivo (químico o biológico) y la calibración, los recipientes y plataformas de reacción, la calidad del agua, el pH y la ionicidad de

2 <https://www.oie.int/es/que-hacemos/normas/codigos-y-manuales/acceso-en-linea-al-manual-terrestre/>

tampones y diluyentes, las temperaturas y duraciones de incubación, y los errores en el rendimiento técnico de la prueba. Se requieren experimentos exhaustivos y bien diseñados para desarrollar y optimizar ensayos con características analíticas favorables. Los principios subyacentes son ampliamente aplicables a todos los tipos de ensayo y, cuando se llevan a cabo con el rigor adecuado, sientan las bases de pruebas diagnósticas de calidad aptas para las finalidades definidas (Bowden et al., 2021). En su revisión de las pruebas de diagnóstico recomendadas por la OMSA, Cullinane y Garvey (2021) concluyeron que el ensayo inmunoenzimático (ELISA) y los ensayos moleculares eran las pruebas recomendadas por la OMSA más utilizadas, razón por la cual los ejemplos de las secciones siguientes se centran en estos métodos (Mayo et al., 2021).

El primer paso en el desarrollo de una prueba es definir la finalidad de la prueba, ya que esto guiará todos los pasos subsiguientes del proceso de validación. Los criterios de validación de la prueba son los rasgos característicos de una prueba que constituyen factores decisivos, mediciones o estándares en los cuales se basa un juicio o decisión. Si tenemos en cuenta las variables que pueden afectar al rendimiento de una prueba, podremos apreciar más claramente los criterios que deben tenerse en cuenta para su validación. Estas variables pueden agruparse en categorías: (a) relativas a la muestra: si son individuales o agrupadas, la composición de la matriz, y las interacciones hospedador/organismo que afecten al analito en cuestión cuantitativa o cualitativamente; (b) relativas al sistema analítico – que incluyen factores físicos, químicos, biológicos y técnicos que afectan a la capacidad de la prueba de detectar en la muestra un analito específico; y (c) relativos a la interpretación del resultado de la prueba – es decir, la capacidad del sistema analítico de predecir de forma precisa el estado del individuo o de la población en relación con la finalidad con la cual se aplica la prueba.

La elección, obtención, preparación, conservación y manejo de las muestras son variables cruciales en el diseño y desarrollo de una prueba para garantizar resultados analíticos válidos. Otras variables, como el transporte, la cadena de custodia, la rastreabilidad de las muestras y el sistema de gestión de la información del laboratorio también son fuentes clave de variación/error que se convierten en especialmente importantes cuando la prueba se utiliza para un análisis sistemático. La integridad de los resultados experimentales obtenidos en el laboratorio durante la ejecución y validación de una prueba será como mucho igual de buena que la calidad de las muestras que se utilicen. Antes de iniciar las tareas de validación de una prueba, es importante prever los factores que pueden afectar negativamente a la calidad de la muestra. Las muestras de referencia que se utilicen en la ejecución y validación de una prueba deberán encontrarse en la misma matriz que se va a utilizar en la prueba (por ejemplo, suero, tejido o sangre total) y deberán ser representativas de la especie que vaya a analizarse mediante dicha prueba. Los materiales de referencia deberán representar adecuadamente el intervalo de concentraciones de analito a detectar mediante la prueba. Los biobancos virtuales se han convertido en recursos relevantes en lo que respecta a reactivos y muestras durante el desarrollo y la validación de pruebas. Por ejemplo, una vez que estuvo disponible la secuencia genómica del coronavirus 2 del síndrome respiratorio agudo severo (SARS-CoV-2), los miembros del *European Virus Archive Global* (EVAg) desarrollaron rápidamente reservas de virus, pruebas de diagnóstico y controles positivos y, en 10 semanas, se habían distribuido más de 1500 productos en todo el mundo con fines de diagnóstico o investigación (Watson et al., 2021). El capítulo 2.2.6 *Elección y uso de muestras y paneles de referencia* ofrece una visión general sobre la elección y el uso de muestras y paneles de referencia para abordar parámetros de validación relevantes, como la repetibilidad, la reproducibilidad, la DSe, la DSp, etc. En los capítulos 1.1.2 y 1.1.3 del *Manual Terrestre* se encuentra información sobre la obtención, preparación, conservación, manejo y transporte de las muestras.

La matriz (suero, heces, tejido, etc.) en la que puede residir el analito buscado puede contener inhibidores endógenos o exógenos que pueden interferir con el rendimiento de la prueba. Esto es especialmente importante en el caso de pruebas que dependen de enzimas, como la reacción en cadena de la polimerasa (PCR) o el ensayo inmunoenzimático (ELISA). Otros factores que también influyen en la concentración y en la composición de un analito buscado (en concreto anticuerpos) en la muestra son en su mayor parte atribuibles al hospedador y pueden ser factores inherentes (como la edad, el sexo, la raza, el estado nutricional, la gestación o la capacidad de respuesta inmunitaria), o bien adquiridos (como la adquisición pasiva de anticuerpos o la inmunidad activa obtenida por la vacunación o la infección). Otros factores que no dependen del hospedador, como la contaminación o el deterioro de la muestra, también pueden llegar a afectar a la capacidad de la prueba de detectar el analito buscado en la muestra. También es importante que los reactivos biológicos estén libres de agentes extraños que, de encontrarse presentes, pudieran conducir a resultados erróneos.

CRITERIOS DE DESARROLLO Y VALIDACIÓN DE PRUEBAS

La exactitud y la precisión son dos parámetros independientes que, en última instancia, definen el rendimiento de una prueba diagnóstica. La sensibilidad, la especificidad y las funciones de estas variables (por ejemplo, los cocientes de probabilidades) cambian en función del punto de corte y representan la exactitud, mientras que la repetibilidad y la reproducibilidad son medidas de la precisión. Para que una prueba sea fiable debe ser exacta y precisa. El rendimiento de la prueba se ve afectado por muchos factores, empezando por la optimización de la prueba. Tras la optimización inicial para una finalidad definida, se comprobarán las características de rendimiento de la prueba (recuadro con los criterios de desarrollo y validación de las pruebas). Es posible que la prueba requiera una optimización adicional o que se observe que es apta para una finalidad definida en base a los resultados del trabajo de validación. La Secretaría para el Registro de Kits de Diagnóstico (SRDK) de la OMSA cuenta con un proceso de validación y certificación de pruebas mediante el cual se pueden registrar las pruebas nuevas que se consideren adecuadas para las finalidades definidas³.

Crterios para el desarrollo y la validación de una prueba

- i) Establecimiento de la/s finalidad/es definida/s
- ii) Optimización
- iii) Estandarización
- iv) Repetibilidad
- v) Sensibilidad analítica
- vi) Especificidad analítica
- vii) Umbrales (puntos de corte)
- viii) Sensibilidad diagnóstica
- ix) Especificidad diagnóstica
- x) Reproducibilidad
- xi) Idoneidad para la/s finalidad/es definida/s

A. FASE DE DESARROLLO DE LA PRUEBA

1. Definición de las finalidades definidas de una prueba

La norma ISO/IEC 17025 (2017) para laboratorios de ensayo y calibración establece que el laboratorio debe utilizar métodos y procedimientos adecuados para todas las actividades de laboratorio (Newberry y Colling, 2021). En otras palabras, el ensayo debe ser "apto para la finalidad". Si no se define a priori la finalidad del ensayo, es probable que se produzcan errores tanto en el diseño de este, como en la determinación de sus parámetros críticos, lo que podría invalidar todo el proceso de desarrollo y validación del ensayo y, en última instancia, dar lugar a un ensayo que no satisfaga las necesidades del usuario. Tales errores pueden producirse, por ejemplo, cuando se selecciona una población de referencia inadecuada que sea heteróloga respecto a la población para la que se está desarrollando la prueba.

Crterios para la interpretación de los resultados

- Valor predictivo positivo (PV+)
- Valor predictivo negativo (PV-)

- Cociente de probabilidades positivo (LR+)
- Cociente de probabilidades negativo (LR-)

La capacidad de evaluación cualitativa y cuantitativa de la capacidad de un resultado positivo o negativo de una prueba, como, por ejemplo, el valor predictivo y el cociente de probabilidades para predecir con un alto nivel de confianza si un animal o población de animales están o no infectados o han estado o no expuestos es el criterio definitivo para la validación de dicha prueba (Sección B.4.2). Esta capacidad depende de que una prueba, desarrollada mediante una cuidadosa optimización y estandarización (apartado A.2.3) ofrezca, mediante la acumulación de datos de validación, confianza en la capacidad de la prueba de rendir de acuerdo con la finalidad definida (Tabla 1).

Con el fin de asegurar que los resultados de la prueba proporcionan útiles inferencias diagnósticas sobre animales o poblaciones de animales respecto a la finalidad definida, el proceso de validación abarca documentación sobre el desarrollo inicial y la realización de la prueba, así como una evaluación continua de los programas de control y garantía de calidad.

La Figura 1 muestra el proceso de validación de la prueba, desde el diseño hasta las fases de desarrollo y validación y la implementación, despliegue y mantenimiento de este.

3 [Registro de kits de diagnóstico - OMSA - Organización Mundial de Sanidad Animal \(woah.org\)](https://www.woah.org/)

El primer paso del desarrollo de una prueba consiste en escoger un tipo de prueba que sea adecuada y que tenga posibilidades de ser validada para un uso concreto (adecuada para la finalidad definida).

Las finalidades más frecuentes de entre las que figuran en el *Manual Terrestre* de la OMSA son las siguientes:

- 1) Contribuir a demostrar la ausencia de infección en una población definida (país/zona/compartimento/rebaño) (prevalencia aparente del 0%):
 - 1a) “Libre” con y/o sin vacunación,
 - 1b) Restablecimiento de la ausencia después de los brotes
- 2) Certificar la ausencia de infección o presencia del agente causal en animales concretos o productos utilizados para el comercio o el transporte.
- 3) Contribuir a la erradicación de la enfermedad o eliminación de la infección en poblaciones definidas.
- 4) Confirmar el diagnóstico de casos clínicos o sospechosos (incluye la confirmación de un resultado positivo en una prueba de cribado).
- 5) Estimar la prevalencia de la infección o exposición para facilitar el análisis de riesgos (encuestas, situación sanitaria del rebaño respecto a la infección, medidas para el control de la enfermedad).
- 6) Determinar el estado inmunitario de animales determinados o de poblaciones (después de la vacunación).

Table 1: Finalidades de las pruebas e importancia proporcional de la sensibilidad diagnóstica (DSe), la especificidad diagnóstica (DSp), el valor predictivo positivo (PV+), el valor predictivo negativo (PV-), el cociente de probabilidades de un resultado positivo en la prueba (LR+) y el cociente de probabilidades de un resultado negativo en la prueba (LR-)

Finalidad	Ejemplos y mediciones de la exactitud diagnóstica en función de la finalidad de la prueba (Reid <i>et al.</i> , 2022)
1a) Mantener el estatus histórico de libre de la enfermedad (con o sin vacunación)	En una población históricamente libre de una enfermedad/agente patógeno concreto, la prevalencia es cero (o cercana a cero). Para ser adecuada para la finalidad definida, la prueba o el algoritmo de prueba debe minimizar la probabilidad de falsos positivos e idealmente requiere una DSp alta, un PV+ alto y un LR+ alto . Esto puede lograrse mediante una sola prueba con una DSp alta o pruebas seriadas ^(*) .
1b) Recuperación del estatus de libre de la enfermedad después de un brote	En el transcurso de un programa exitoso de control de una enfermedad, se puede esperar un cambio gradual en la prevalencia, de alta (durante el pico del brote) a baja (al final del brote). Durante las primeras fases de un programa de pruebas destinadas a demostrar la ausencia de una enfermedad, cuando la prevalencia de la enfermedad se mantiene en niveles no insignificantes, para que la prueba sea adecuada para su finalidad necesita una DSe alta, un PV alto y un LR alto . Este método minimiza la probabilidad de falsos negativos y permite la detección de individuos positivos. Esto puede lograrse mediante una única prueba con una DSe alta o con pruebas en paralelo ^(*) . Al final del programa de control de una enfermedad, cuando se hayan eliminado de la población los animales infectados restantes, la prevalencia de la enfermedad será muy baja, por lo que probablemente habrá que modificar el algoritmo de las pruebas de detección para aumentar la DSp (y mejorar así el PV+ y el LR+) de forma similar a como se hace en 1a.
2) Certificar el estatus de libre de la infección o del agente patógeno en animales o productos determinados con fines de comercio/desplazamientos	A efectos del comercio y los desplazamientos, es necesario minimizar la probabilidad de falsos negativos. De lo contrario, animales infectados podrían ser objeto de comercio o traslado, con la posibilidad de propagar la infección a poblaciones sanas no infectadas. Dado que la prueba se aplica a individuos, no se dispone de información, o ésta es escasa, sobre la prevalencia o la probabilidad de infección antes de la prueba. Para que la prueba o el algoritmo de la prueba sean adecuados para su finalidad, deben minimizar la probabilidad de falsos negativos y, en condiciones ideales, deben tener una DSe, un PV- y un LR- altos . Esto puede lograrse mediante una única prueba con una DSe alta o con pruebas en paralelo ^(*) .
3) Contribuir a la erradicación de la enfermedad o a la eliminación de la infección de poblaciones definidas	Esta finalidad sigue un patrón similar al de 1b, donde se espera que la prevalencia disminuya de alta a baja a lo largo del tiempo en una población definida
4) Confirmar el diagnóstico de casos sospechosos o clínicos (incluye la confirmación de positivos con pruebas de cribado)	El objetivo de una prueba confirmativa es minimizar la probabilidad de un falso positivo. <i>Confirmación de casos clínicos</i> A efectos de confirmación de un caso clínico, lo ideal es una prueba con una DSp alta, un PV+ alto y un LR+ alto . Debido a la manifestación clínica de la enfermedad y a la alta carga patógena esperada, la DSe no se considera tan relevante. <i>Confirmación positivos con pruebas de cribado</i> Las pruebas de cribado se aplican a poblaciones sanas. Suelen tener una DSe alta para garantizar que no se pasen por alto individuos infectados. Se considera que el animal es positivo sólo si se confirma mediante una prueba de confirmación con una DSp alta. En este caso, la prueba confirmativa* debe tener una DSp alta, un PV+ alto y un LR+ alto . Este método sigue el algoritmo de pruebas seriadas ^(*) .
5) Estimar la prevalencia de infección o exposición para facilitar el análisis del riesgo	Los epidemiólogos necesitan estimaciones fiables de la precisión de las pruebas para diseñar planes de muestreo para estudios de prevalencia, encuestas, determinaciones de la situación sanitaria de los rebaños respecto a la infección y decisiones acerca de medidas de control de enfermedades. El uso de una prueba de cribado con una DSe alta seguida de una prueba confirmativa con una DSp alta es un método común para las estimaciones de prevalencia.
6) Determinar el estatus inmunitario respecto a una enfermedad a) en animales individuales post-vacunación b) estimar la seroprevalencia post- vacunación (investigación y seguimiento de la eficacia vacunal)	Para esta finalidad, el objetivo es tener una DSp, un PV+ y un LR+ altos . Un falso positivo podría tener consecuencias fatales, ya que dicho animal podría, de hecho, no estar vacunado/protegido. Cuanto mayor sea la precisión de la prueba, más exacta será la estimación de la seroconversión postvacunal en individuos y poblaciones. Un ejemplo es la prueba de neutralización vírica con inmunofluorescencia indirecta (FAVN, por las siglas en inglés de <i>fluorescent antibody virus neutralisation</i>), que se usa para evaluar el estatus inmunitario de perros y gatos tras la vacunación contra el virus de la rabia. Para los desplazamientos internacionales, se considera que un resultado > 0,5 UI/ml representa una protección aceptable.

(*)Pruebas múltiples

Aplicar pruebas múltiples significa utilizar más de una prueba para determinar si un animal está o no infectado. Los algoritmos más comunes son las pruebas seriadas o en paralelo. Por ejemplo, si se utilizan dos pruebas en serie, una muestra se considera positiva sólo si dan positivo tanto la primera prueba como la segunda. Las pruebas en serie aumentan la DSp pero disminuyen la DSe, y aumentan tanto el PV+ como el LR+. Las pruebas confirmativas siguen el método de las pruebas en serie porque un resultado positivo en una prueba de cribado (DSe alta) debe confirmarse mediante una segunda prueba que tenga una DSp alta. La prueba confirmativa debe tener al menos la misma DSe que la prueba de cribado, de lo contrario podría generar falsos negativos, que en este algoritmo se considerarían verdaderos negativos. Las pruebas confirmativas aumentan la DSp pero disminuyen la DSe, y aumentan tanto el PV+ como el LR+. Si se utilizan dos pruebas en paralelo, una muestra se considera positiva si cualquiera de las pruebas o ambas dan positivo. Las pruebas paralelas aumentan la DSe pero disminuyen la DSp, y aumentan tanto el PV- como el LR-. Además, para que los algoritmos de pruebas múltiples sean eficaces, las pruebas de cribado y de confirmación no deben ser condicionalmente dependientes (por ejemplo, deben medir el mismo analito, como en el caso de dos ELISA que utilicen el mismo antígeno). Una dependencia positiva en la sensibilidad de la prueba reduce la sensibilidad de la interpretación de las pruebas en paralelo, mientras que una dependencia positiva en la especificidad de la prueba reduce la especificidad de la interpretación de las pruebas seriadas (Gardner et al., 2000).

2. Desarrollo de una prueba – estudios experimentales

2.1. Diseño y demostración del método analítico

Para el diseño de todos los pasos de una nueva prueba todavía no validada, o de una prueba existente que se esté modificando, son necesarios un conocimiento previo, así como una buena reflexión y planificación. Se ofrece orientación respecto a las recomendaciones para la validación de las pruebas de diagnóstico⁴, que cubren las mejores prácticas para el desarrollo y validación de las pruebas destinadas a la detección de distintos analitos, como anticuerpos, antígenos o ácido nucleico, Capítulos 2.2.1 *Desarrollo y optimización de pruebas de detección de anticuerpos*, 2.2.2 *Desarrollo y optimización de pruebas de detección de antígenos*, y 2.2.3 *Desarrollo y optimización de pruebas de detección de ácidos nucleicos*, respectivamente.

El desarrollo de todas las pruebas depende de las muestras de referencia del analito, que representan el analito buscado, la matriz en la que se encuentra el mismo y la población a la que se destina la prueba. Las muestras de referencia pueden ser sueros, líquidos (incluidos jugos de carne) o tejidos que contengan el analito de interés o una estructura genómica que concuerde con el analito buscado. Estos materiales de referencia se utilizan en experimentos llevados a cabo durante todo el proceso de desarrollo y también en la validación de la prueba.

Los factores que afectan a las características analíticas de las pruebas de diagnóstico son numerosos y pueden variar según cada tipo de prueba; por ejemplo, los principales factores que afectan a las características analíticas de las pruebas serológicas y moleculares se describen en Bowden et al. (2021).

En el caso de las pruebas moleculares, la DSe depende más de la capacidad de obtener el analito diana en una muestra procesada de un animal que tiene la enfermedad, que de la capacidad inherente del ensayo para detectar concentraciones de analito muy bajas. Aunque un ensayo puede ser extremadamente sensible desde el punto de vista analítico, como la PCR en tiempo real, esto no siempre se traduce en una DSe alta, debido a las posibles deficiencias del muestreo (volumen pequeño, sustancias inhibidoras, variaciones en el espectro clínico de la enfermedad en un animal individual o eficiencia de la extracción). La inhibición de la Taq polimerasa debida a las características de la muestra puede ocasionar falsos negativos en una prueba que, por lo demás, tenga una sensibilidad analítica alta (ASe). Además, al evaluar dos PCR en tiempo real independientes para el virus de la peste porcina africana (VPPA), se observó que ambas tenían una ASe comparable (equivalente a 14 copias genómicas del VPPA) cuando se utilizaba una construcción plasmídica de alta calidad que contenía el VP72 del VPPA como molde. Sin embargo, al evaluar varias muestras obtenidas de cerdos enfermos clínicamente,

4 https://www.woah.org/fileadmin/Home/esp/Health_standards/tahm/2.02.00_INTRODUCTION.pdf

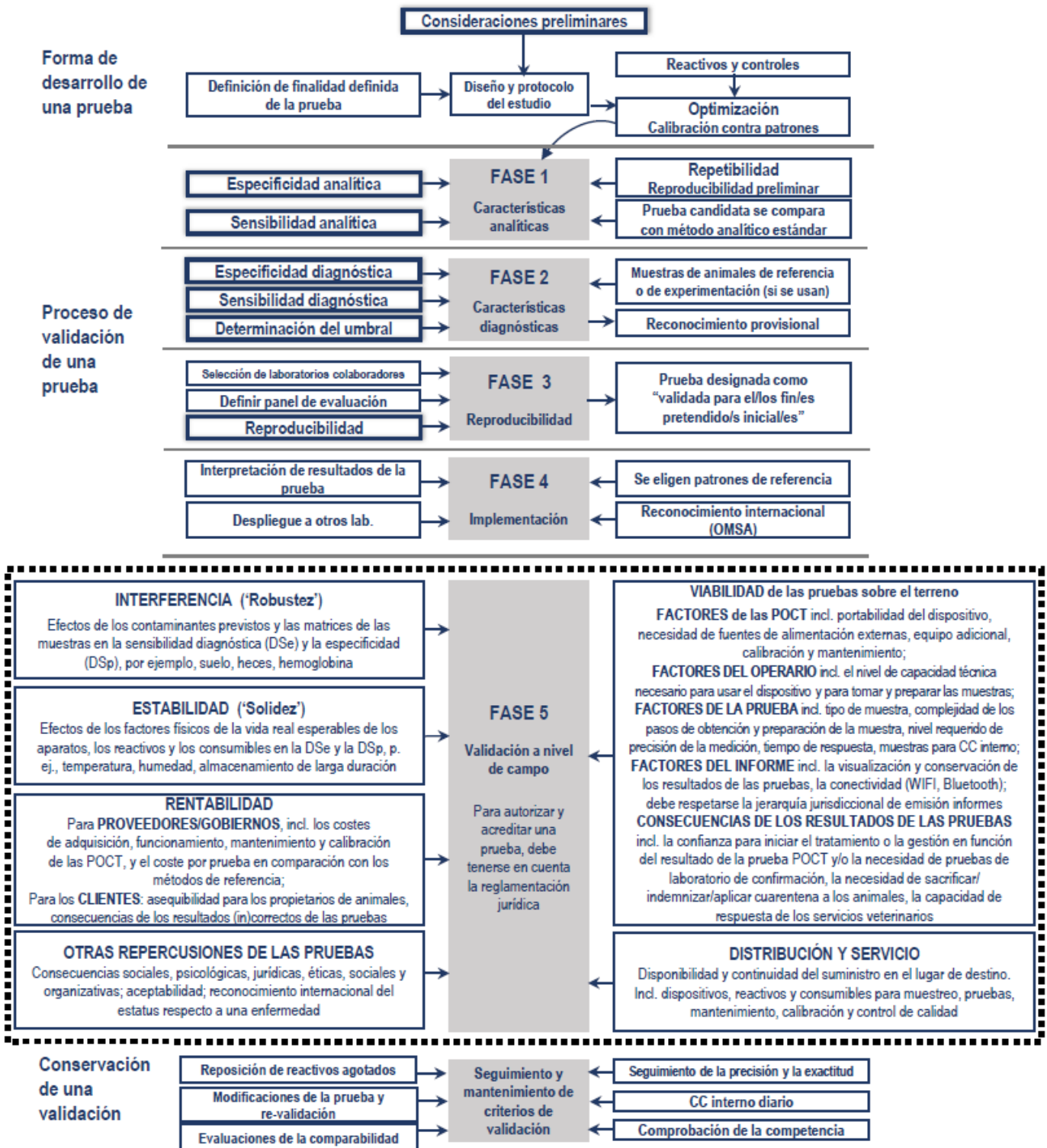


Figura 1. Modificación de las fases generales de desarrollo y validación de pruebas (Fases 1–4 y conservación de la validación) con los criterios de validación destacados en negrita en los recuadros sombreados. Los criterios específicos para la validación a nivel de campo de las POCT se añaden en la Fase 5 dentro del recuadro punteado (Halpin et al., 2021); (CC = control de calidad).

se determinó que la DSe resultante era del 83% para un ensayo y del 92% para el otro (datos no mostrados). El tamaño del amplicón del ensayo con la DSe más baja (250 pares de bases o pb) fue significativamente mayor que el del otro ensayo (75 pb). En general, es esperable que la eficacia de la amplificación disminuya al aumentar el tamaño del amplicón. Esta característica se acentúa cuando se analizan ácidos nucleicos víricos extraídos de muestras clínicas, cuya calidad puede verse afectada por

diversos grados de degradación. Este resultado no se evidenció cuando se utilizó el plásmido como molde durante la determinación de la ASe del ensayo. Además, como un ensayo como la PCR en tiempo real tiene una sensibilidad analítica tan alta, se debe tener mucho cuidado para evitar la contaminación por arrastre con el molde amplificado previamente. Dicha contaminación causaría un falso positivo en un ensayo que, por lo demás, también se considera que tiene una sensibilidad analítica alta (Bowden y Wang 2021).

2.2. Estandarización y optimización

La optimización es el proceso mediante el cual la mayor parte de parámetros físicos, químicos y biológicos de una prueba se evalúan y se ajustan para garantizar que las características de rendimiento de la prueba se adapten mejor a la aplicación pretendida. Es útil seleccionar por lo menos tres muestras de referencia bien definidas, que contengan el analito abarcando resultados desde positivo alto a negativo (por ejemplo, positivo fuerte, positivo débil y negativo). En teoría, las muestras deberían representar tanto a los animales que se consideran infectados como a los no infectados dentro de la población que finalmente va a ser objeto de la prueba. Sin embargo, no siempre es posible obtener estas muestras de referencia, sobre todo en el caso de las pruebas de detección de ácido nucleico y antígeno. La alternativa de preparar muestras de referencia a las que se hayan añadido agentes patógenos de cultivo o sueros positivos no es tan adecuada, puesto que estas muestras no representan exactamente la interacción entre matriz y agente que tiene lugar en condiciones naturales (véase también el Capítulo 2.2.6). Cuando no existe otra alternativa, tal vez la única posibilidad sea añadir a la muestra una cantidad conocida del analito o agente derivado del cultivo, o bien diluir un suero positivo alto en suero negativo de la misma especie. En cualquier caso, es imprescindible que la matriz a la que se añade o en la que se diluye el analito sea idéntica o se parezca lo máximo posible a la de las muestras que finalmente se analizarán mediante la prueba. Lo ideal es que las muestras de referencia estén bien caracterizadas por una o, preferiblemente, al menos dos metodologías alternas. Estas muestras pueden utilizarse en experimentos para determinar si la prueba es capaz de distinguir diferentes cantidades del analito, distinguir los analitos buscados de otros que estén estrechamente relacionados con ellos, y optimizar las concentraciones de reactivos y perfeccionar el protocolo. En principio, para todos los tipos de prueba es muy conveniente preparar y almacenar una cantidad suficiente de cada muestra de referencia en alícuotas para su uso en cada ejecución de la prueba candidata, ya que se evalúa a lo largo de todo el proceso de desarrollo y validación. Cambiar las muestras de referencia durante el proceso de validación introduce una variable difícil de resolver que puede socavar gravemente la interpretación de los datos experimentales y, por tanto, la integridad del proceso de desarrollo y validación.

El laborioso proceso de optimización de una prueba es fundamental y crítico para lograr un rendimiento analítico fiable y predecible. La evaluación científica y la aplicación de las mejores prácticas científicas, como se indican en Bowden *et al.* (2021), se recomiendan para guiar la optimización de todos los elementos del desarrollo y validación de una prueba. El enfoque descrito proporciona una base firme para el desarrollo de una prueba fiable. A menudo se han desarrollado prototipos de pruebas utilizando los reactivos y el equipo que se tenían a mano en el laboratorio. Sin embargo, si la prueba ha sido diseñada para su ser utilizada con fines diagnósticos en múltiples laboratorios, la estandarización se vuelve crucial. Deben describirse detalladamente todas las formulaciones de sustancias químicas y tampones. Todos los reactivos deben ser definidos con respecto a la pureza y al grado (incluida el agua). Debe establecerse y documentarse cuáles son los intervalos de trabajo aceptables para parámetros como el pH, la molaridad, etc. De igual forma, para los productos biológicos también deben definirse los estándares de calidad, pureza, concentración y reactividad. Tanto en productos químicos como en productos biológicos también deben tenerse en cuenta los periodos de validez y las condiciones de almacenamiento. También deben establecerse los intervalos aceptables de los tiempos de reacción y las temperaturas. Debe describirse detalladamente el equipo esencial para optimizar el rendimiento de la prueba, incluyendo especificaciones de funcionamiento y de calibración. El control del proceso (calidad) debe formar parte de la optimización y debe tenerse en cuenta desde el principio, y no al final del desarrollo de la prueba, como suele ocurrir. Además de lo anterior, otros aspectos posteriores, como la obtención, manipulación e interpretación de los datos también pueden requerir estandarización y optimización. Finalmente, todos estos parámetros, una vez optimizados, deben describirse en detalle en el protocolo del método analítico.

Durante la optimización de la una prueba, es importante tomar nota de los pasos del procedimiento y de los parámetros de la prueba cuyo intervalo de valores en el que la prueba ofrece un rendimiento óptimo es estrecho, ya que son puntos críticos que en último término afectan a la fiabilidad de la prueba (véase

el apartado A.2.7). En el caso de ciertos tipos de prueba, determinados pasos del procedimiento pueden influir más que otros en el rendimiento final de la prueba (véase el apartado B.5 abajo y el Capítulo 2.2.8 *Comparabilidad entre pruebas tras realizar cambios en un método analítico validado* para información adicional sobre cómo establecer la comparabilidad cuando se cambian reactivos o procesos; y también Reising *et al.*, 2021).

En los siguientes apartados se presentan varias muestras de referencia de analitos y otros controles del proceso que se incluyen sistemáticamente en cualquier sistema analítico. Estos proporcionan las funciones críticas de seguimiento de la prueba que requieren atención especial durante la optimización de este. Además, para asegurar la estabilidad tienen que garantizarse la preparación y el almacenamiento adecuados de todos los reactivos biológicos y materiales de referencia (véase el capítulo 1.1.2; Watson *et al.*, 2021).

2.3. Intervalo de funcionamiento de la prueba

El intervalo de funcionamiento de una prueba es el intervalo de concentraciones o títulos del analito en el cual el método aporta una exactitud y precisión idóneas. La exactitud es la cercanía de un valor analítico al valor esperado (real) (media o mediana) para un reactivo estándar de referencia de concentración o título conocidos. La precisión es el grado de dispersión (varianza, desviación estándar [DE] o coeficiente de variación [CV]) dentro de una serie de mediciones de la misma muestra analizada en unas condiciones especificadas. Las fuentes de variación en el laboratorio que afectan a la precisión del ensayo son: 1) las que generan variación dentro de un mismo ensayo, 2) las que generan variación entre ensayos simultáneos, 3a) las que generan variación entre ensayos realizados a distintas horas del mismo día o en días diferentes en condiciones similares, 3b) las que generan variación entre ensayos realizados en días diferentes con operarios diferentes, 4) las que generan variación entre laboratorios. En este capítulo, las categorías 1-3 son estimaciones de la repetibilidad, y la categoría 4 es sinónimo de reproducibilidad. Es probable que la repetibilidad de los resultados de operarios de laboratorios diferentes sea menor (mayor DE, mayor CV) que la de los resultados de operarios que trabajan en un mismo laboratorio.

Durante el desarrollo de la prueba, se determinan los límites inferior y superior del intervalo de funcionamiento. Para determinar formalmente este intervalo, se selecciona una muestra de referencia positiva alta (lo ideal es que esta muestra se encuentre entre las tres muestras descritas en Apartado A.2.3). Se lleva a cabo una dilución seriada de esta muestra positiva alta hasta la extinción de respuesta de la prueba en una matriz sin analito que tenga la misma constitución que la matriz de muestra de los animales de la población a la que se destina la prueba. Los resultados se trazan en una "curva de respuesta", en la que la respuesta (por ejemplo, la densidad óptica, el ciclo umbral, los recuentos de agentes patógenos, etc.) es función de la concentración (cantidad) de analito. La curva establece el intervalo de funcionamiento de la prueba. Si se observa que el intervalo es inaceptable para la finalidad definida, tal vez sea necesaria una mayor optimización. La curva de calibración típica para la mayoría de las pruebas tiene una forma sigmoidea. Los datos se transforman para aproximar una relación lineal entre la respuesta y la concentración empleando un algoritmo adecuado (Findlay y Dillard, 2007).

2.4. Factores inhibidores en la matriz de la muestra

Cada matriz que vaya a utilizarse en una prueba debe emplearse en el proceso de validación. Algunas matrices de muestra incluyen factores inhibidores que interfieren con el rendimiento de determinados tipos de prueba. El suero, en concreto si está hemolizado, puede contener factores tóxicos para las células utilizadas en las pruebas de neutralización de virus, mientras que las sustancias endógenas que se encuentran en algunos tejidos y líquidos pueden interferir con las pruebas basadas en la fijación de ligandos y en enzimas, como el ELISA, o bien inhibirlas. Las muestras de heces, tejidos autolisados y semen tienden a contener más sustancias que interfieren y, por lo tanto, son más problemáticas para el rendimiento analítico que el suero, la sangre o los tejidos frescos. En el caso de los ensayos moleculares, en la matriz puede haber inhibidores, interferentes y degradantes de las enzimas de la mezcla de reacción. La selectividad es esencialmente la prueba de detección en presencia de inhibidores. La variación de la matriz de la muestra es una de las fuentes de error más importantes en las mediciones analíticas, aunque puede ser una de las menos reconocidas. En la evaluación de la ASp, deben utilizarse matrices adecuadas para la finalidad definida, como tejidos sólidos, sangre total o hisopos. Las sustancias que interfieren (inhibidores) pueden proceder de las muestras, como, por ejemplo, la hemoglobina (Schrader *et al.*, 2012; Wilson, 1997), o del entorno o del proceso de toma y transporte de

las muestras (por ejemplo, medios y anticoagulantes) (Druce *et al.*, 2012; Garcia *et al.*, 2002; Gibb *et al.*, 1998; Miyachi *et al.*, 1998; Yokota *et al.*, 1999;). Por ejemplo, la sangre total es una de las muestras más importantes y habituales en la RT-PCR en tiempo real para Hendravirus (HeV). Se descubrió que la heparina de litio, un anticoagulante de uso habitual, tenía un impacto significativo en las RT-PCR en tiempo real para HeV, puesto que producía falsos negativos. En cambio, el anticoagulante ácido etilendiaminotetraacético (EDTA) tenía un efecto inhibitorio comparativamente bajo. Además, en la RT-PCR en tiempo real para el virus de la lengua azul (VLA), diluir la sangre 1/10 (vol/vol) en solución salina tamponada con fosfato dio lugar a una mayor sensibilidad, evidenciada por valores de Ct (ciclo umbral) más bajos, en comparación con la sangre sin diluir. Una revisión reciente de los diferentes tipos de controles internos existentes para monitorizar la inhibición de la PCR en tiempo real (Yan *et al.*, 2020) proporciona información sobre las estrategias de control interno como componente rutinario de gestión de calidad en las pruebas moleculares de la veterinaria. Los datos recogidos durante el proceso de validación relativos al rendimiento del ensayo cuando se utilizan matrices de muestras objeto análisis permitirán tomar una decisión basada en el riesgo acerca de si debe incluirse un control de inhibición para cada muestra o si es improbable que el sistema analítico se vea afectado por la inhibición. Si las sustancias inhibitorias constituyen un problema importante, deberá incluirse un control de inhibición para cada muestra a analizar (Sección 1.2; Bowden *et al.*, 2021).

2.5. Robustez

La robustez es la capacidad de una prueba de no resultar afectado por pequeñas variaciones en las situaciones en que se lleva a cabo el análisis, que podrían tener lugar en el curso de la prueba. La evaluación de la robustez debe comenzar durante las etapas de desarrollo y optimización de la prueba. Las variaciones intencionadas en los parámetros del método pueden abordarse en los experimentos una vez establecidas las condiciones óptimas de una prueba. Sin embargo, cuando se utilizan valoraciones multifactoriales de los reactivos para optimizar la prueba, pueden aparecer indicios de problemas en la robustez. Si pequeñas diferencias en las condiciones o en las concentraciones de reactivo causan una variabilidad inaceptable, lo más probable es que la prueba no sea robusta. Conocer cuanto antes esta situación constituye un punto crítico de la toma de decisiones para determinar si vale la pena seguir con la validación de la prueba, porque si esta no es robusta en un laboratorio en condiciones bastante ideales, es improbable que sea reproducible cuando se transfiera a otros laboratorios.

Los factores que con mayor probabilidad afectarán a la robustez de la prueba son el pH, la temperatura, el lote de los reactivos o la marca de las placas de microtitulación y factores relacionados con la matriz acuosa u orgánica (Bowden *et al.*, 2021; Dejaegher y Vander Heyden, 2006). Una vez se ha terminado la optimización, la robustez de la prueba deviene parte de la evaluación de la repetibilidad. En el caso de las pruebas a pie de establo (POCT) (véase la Figura 1, Fase 5), también es necesario evaluar la robustez (expresada como la ausencia de influencia de las variables operativas y ambientales del método analítico en los resultados de las pruebas), ya que el rendimiento de las POCT no debe verse afectado fácilmente por la pericia del operario, las fluctuaciones de temperatura, la humedad, la luz solar u otros factores ambientales.

2.6. Calibración de la prueba frente a reactivos estándar

2.6.1. Estándares de referencia a nivel internacional y nacional

Lo ideal es que los estándares de referencia a nivel internacional de la OMSA y otros, que contienen una concentración o título conocidos del analito, sean los reactivos frente a los cuales todas las pruebas están estandarizadas (véanse las Directrices de la OMSA⁵ y también el Capítulo 2.2.6). Estos estándares los preparan y distribuyen Laboratorios de Referencia de la OMSA y otros laboratorios de referencia a nivel internacional. Los estándares de referencia a nivel nacional se calibran por comparación con un estándar de referencia a nivel internacional siempre que ello es posible, y los prepara y distribuye un laboratorio de referencia a nivel nacional. En ausencia de un estándar de referencia a nivel internacional, un estándar de referencia a nivel nacional se convierte en el estándar de comparación para la prueba candidata. Estos estándares se caracterizan con gran detalle mediante un exhaustivo análisis, y es preferible optar por métodos

5 <http://www.oie.int/es/nuestra-experiencia-cientifica/productos-veterinarios/reactivos-de-referencia/>

de caracterización, preparación y almacenamiento que hayan sido publicados en publicaciones revisadas por expertos (Watson et al., 2021).

2.6.2. Estándar interno

Un estándar de referencia interno en general debe calibrarse frente a un estándar Internacional o Nacional. En ausencia de cualquiera de los calibradores y en la medida de lo posible, el estándar interno se caracteriza con gran detalle de igual forma que los estándares internacionales y nacionales (Capítulo 2.2.6). Este estándar interno local, por tanto, se convierte en el mejor estándar disponible, y se mantiene en volúmenes de alícuotas suficientes para un uso periódico como el estándar frente al cual se deberán calibrar los estándares de trabajo.

2.6.3. Estándar de trabajo

Se calibran uno o más estándares de trabajo, a menudo denominados controles del analito o del proceso, frente a un estándar internacional, nacional o interno, se preparan en grandes cantidades, en alícuotas, y se guardan para poder utilizarlos cada vez que se ejecute la prueba con fines de diagnóstico.

2.7. “Normalización” de los resultados de la prueba frente a un estándar de trabajo

Debido a la variación inherente en los resultados brutos de la prueba que se observan con frecuencia entre las ejecuciones de una misma prueba o entre laboratorios utilizando pruebas iguales o similares, es casi imposible comparar directamente datos (semi-)cuantitativos. Para mejorar visiblemente la comparabilidad de los resultados de la prueba, tanto intra como entre los laboratorios, se utiliza uno o más reactivo/s estándar de trabajo en cada ejecución de la prueba. A continuación, los valores brutos de la prueba para cada muestra problema pueden convertirse en unidades de actividad respecto al estándar/es de trabajo mediante un proceso denominado “normalización”. Los valores “normalizados” se pueden expresar de muchas maneras, como en porcentaje de un control positivo (por ejemplo, en un ELISA), o como la concentración estimada, por ejemplo, copias genómicas o título de un analito derivado de una curva estándar, como un ciclo umbral (Ct, por sus siglas en inglés) en una prueba con sonda de hidrólisis. Es una buena práctica el incluir estándares de trabajo en todas las ejecuciones de la prueba durante el desarrollo y validación de la prueba, ya que esto permite una “normalización” de los datos, lo cual proporciona un medio válido para la comparación directa de los resultados entre las ejecuciones de una prueba. Es indispensable controlar la variación (absoluta) de los estándares de normalización, puesto que de lo contrario la normalización puede introducir un sesgo. Véase los Capítulos 2.2.1 *Desarrollo y optimización de las pruebas de detección de anticuerpos*, 2.2.2 *Desarrollo y optimización de las pruebas de detección de antígeno* y 2.2.3 *Desarrollo y optimización de las pruebas de detección de ácido nucleico*, para más información; Bowden et al. (2021).

2.8. Estudio provisional de la repetibilidad

La evaluación de la repetibilidad debe empezar durante las fases de desarrollo y optimización de la prueba. Conocer inicialmente esta situación resulta fundamental para determinar si vale o no la pena seguir adelante con la validación de la prueba.

La repetibilidad se vuelve a verificar durante la Fase 1 de la validación de la prueba (apartado B.1.1). Cuando la prueba optimizada se ejecuta en un laboratorio de rutina o en condiciones de campo (Fase 4 de la validación de la prueba), se realiza un seguimiento continuo de la repetibilidad como parte de los procedimientos de control del proceso a lo largo de toda la prueba (véase el apartado B.5.1).

B. FASE DE VALIDACIÓN DE LA PRUEBA

La “validación” es un proceso que determina la idoneidad de una prueba que se ha desarrollado, optimizado y estandarizado adecuadamente para una/s finalidad/es definida/s. La validación incluye estimaciones de las características de rendimiento analítico y diagnóstico de una prueba. En el contexto de este documento, una prueba que ha superado las tres primeras fases de la validación (Figura 1), incluida la caracterización del rendimiento, puede designarse como “validada para la/s finalidad/es inicial/es definida/s”. La falta de un número estadísticamente robusto de muestras de animales infectados y no infectados se observa con frecuencia en los

casos de enfermedades zoonóticas nuevas y emergentes, y puede ser un obstáculo importante para obtener estimaciones fiables de la precisión (Colling *et al.*, 2018; Stevenson *et al.*, 2021). En estas circunstancias, las autoridades nacionales o los socios comerciales pueden reconocer provisionalmente las pruebas con sensibilidad (ASe) y especificidad (ASp) analíticas aceptables y con resultados de repetibilidad y de DSe y DSp prometedoros pero preliminares basados en un número bajo de muestras, hasta que los resultados de pruebas adicionales confirmen la idoneidad general para la finalidad definida. Muestras bien caracterizadas se han utilizado con éxito para diferentes finalidades en estudios de comparación entre laboratorios para evaluar la idoneidad para una finalidad definida (capítulo 2.2.6; Gardner *et al.*, 2021). Las nuevas plataformas que pueden detectar múltiples agentes patógenos simultáneamente, como, por ejemplo, las tecnologías multiplex, la secuenciación de alto rendimiento, los ensayos con biomarcadores o las pruebas a pie de establo representan nuevos retos para los estudios de validación orientados a finalidades definidas (Bath *et al.*, 2020; Halpin *et al.*, 2021; Reid *et al.*, 2021; van Borm *et al.*, 2016).

1. Fase 1 – Las características del rendimiento analítico

Lo ideal es que el diseño de los estudios descritos en los siguientes apartados se lleve a cabo con la ayuda de un estadístico y de un experto en enfermedades para asegurarse de que el tamaño de la muestra y el enfoque experimental sean válidos. Es posible diseñar experimentos que de manera eficiente proporcionen información sobre fuentes probables de variación en la precisión de la prueba intra e interlaboratorio⁶, lo cual definirá las características de rendimiento de la prueba. La elección de los microorganismos, las cepas o los serotipos para evaluar la sensibilidad y la especificidad analíticas deberá reflejar el conocimiento actual y, por lo tanto, informar sobre cuál es el mejor diseño experimental posible para detectar analitos determinados.

1.1. Repetibilidad

La repetibilidad es el nivel de concordancia entre resultados de réplicas de una muestra tanto intra como entre ejecuciones del mismo método analítico en un laboratorio determinado. La repetibilidad se calcula mediante la evaluación de la variación en los resultados de varias réplicas. El número de réplicas debe determinarse preferiblemente consultando a un estadístico, y se sugiere un mínimo de tres muestras que contengan una actividad del analito que se sitúe en el intervalo de funcionamiento de la prueba. A continuación, de cada una de estas muestras se toman alícuotas y se depositan en un recipiente individual como réplicas idénticas de la muestra original que contienen las concentraciones originales de analito y matriz (véase el Capítulo 2.2.6). Después, cada réplica se procesa pasando por todos los pasos de la prueba, incluidos la creación de la dilución de trabajo, como si fuera una muestra problema procedente de la población objeto de la prueba. No resulta aceptable preparar una dilución final de trabajo de una muestra en un solo tubo a partir del cual se pipeteen alícuotas diluidas a vasos de reacción, ni crear réplicas de una extracción de ácido nucleico en lugar de extraer cada réplica antes de llevar a cabo la dilución en el interior de los vasos de reacción. Estas “muestras” no constituyen réplicas válidas para los estudios de repetibilidad. Se analiza la variación entre diferentes ejecuciones de la misma prueba utilizando las mismas muestras en múltiples ejecuciones llevadas a cabo por dos o más técnicos, preferiblemente en varios días. Bowden y Wang (2021) proporcionan un ejemplo de repetibilidad entre ensayos usando tres PCR en tiempo real para el virus de la peste porcina africana; estas pruebas se evaluaron utilizando una muestra débilmente positiva para este virus. Las pruebas, incluida la extracción del ADN vírico, se ejecutaron en diez días diferentes por parte de distintos operarios y en el mismo laboratorio. La variación en los resultados de las distintas réplicas se puede expresar como desviaciones estándar, coeficientes de variación (desviación estándar ÷ media de réplicas), u otras opciones posibles (véase el Capítulo 2.2.4 *Incertidumbre de la medición* para las evaluaciones de la repetibilidad).

1.2. Especificidad analítica

La especificidad analítica (ASp) es la capacidad de la prueba de distinguir entre el analito buscado (como un anticuerpo, microorganismo o secuencia genómica) y analitos no buscados, incluidos componentes

6 La precisión puede evaluarse de varias maneras analizando la misma muestra replicada: 1) dentro de una placa o placas en una ejecución del ensayo, 2) entre placas analizadas simultáneamente dentro de una ejecución del ensayo, 3a) entre ejecuciones del ensayo a diferentes horas del mismo día o en días diferentes en condiciones similares, 3b) entre ejecuciones del ensayo en días diferentes por parte de operarios diferentes, 4) entre laboratorios. En este capítulo, las categorías de precisión 1-3 son estimaciones de la repetibilidad, y la categoría de precisión 4 es sinónimo de reproducibilidad. Los niveles 3a y 3b también se conocen como precisión intermedia.

de la matriz. La evaluación es cualitativa y la elección y procedencias de los tipos de muestra, microorganismos y secuencias escogidos para evaluar la ASp deben reflejar la finalidad de la prueba y el tipo de prueba. En los Capítulos 2.2.1, 2.2.2 y 2.2.3 se ofrecen información sobre pruebas de anticuerpos, antígenos y ácidos nucleicos, respectivamente. Por ejemplo, la especificidad analítica de una PCR para la fiebre aftosa podría evaluarse analizando muestras de referencia que se sepa que son positivas para el virus de la estomatitis vesicular, la enfermedad vesicular porcina y/o la fiebre catarral maligna. Las evaluaciones de la especificidad analítica no pueden determinar toda la gama de posibles analitos que generen reacción cruzada presentes en la población, ni tener en cuenta la toda la variabilidad de muestras a nivel de población; por lo tanto, la determinación de la especificidad analítica no es un sustituto de la evaluación de la especificidad diagnóstica. La ASp se documenta durante la validación de la Fase 1, y se identifican las reacciones cruzadas (Ludi *et al.*, 2021). La reactividad cruzada (ASp inferior al 100%) puede ser aceptable en función del uso propuesto de la prueba. La influencia de la reactividad cruzada se documenta en mayor medida durante la Fase 2 (establecimiento de la DSP) y se evalúa en el momento de implementar la Fase 4.

1.2.1. Selectividad

La selectividad es el grado en que un método puede cuantificar con exactitud el analito buscado en presencia de: 1) interferentes, tales como componentes de la matriz (por ejemplo, inhibidores de enzimas en la mezcla de reacción); 2) productos de degradación (por ejemplo, factores tóxicos); 3) una unión no específica de los reactivos a una fase sólida (por ejemplo, conjugado de un ELISA adsorbido a un pocillo de la placa de microtitulación); y 4) anticuerpos contra la vacunación, que pueden confundirse con anticuerpos contra la infección activa. Estos interferentes pueden causar falsas reducciones o incrementos en las respuestas de la prueba que afectarán negativamente a su especificidad analítica. Vessman, *et al* (2001) aportaron una útil visión general de la selectividad según se define para la química analítica, de la cual se dedujo una modificación aquí descrita para la aplicación a las pruebas diagnósticas.

1.2.2. Exclusividad

La exclusividad es la capacidad del método analítico de detectar un analito o secuencia genómica que es propia del microorganismo buscado, y excluye todos los demás microorganismos conocidos que pudieran dar una reacción cruzada. Esto también definiría un ensayo confirmativo. Por ejemplo, un ensayo para detectar subtipos H5 del virus de la influenza aviar (VIA) debe evaluarse para detectar posibles reacciones cruzadas con subtipos del VIA que no sean H5. Las pruebas de especificidad también deben incluir otros microorganismos que causen signos clínicos similares, con el fin de demostrar la utilidad del ensayo para la detección diferencial del microorganismo diana.

1.2.3. Inclusividad

La inclusividad es la capacidad de una prueba de detectar varias cepas o serovariedades de una especie, varias especies de un género o una agrupación similar de microorganismos o anticuerpos estrechamente emparentados. Caracteriza el ámbito de aplicación de una prueba de cribado, e por ejemplo, un ELISA específico para el grupo de la lengua azul (VLA) que detecte anticuerpos contra todos los serotipos del VLA o un ELISA para las PNE de la fiebre aftosa que detecte anticuerpos contra los siete serotipos de la fiebre aftosa.

1.3. Sensibilidad analítica

El límite de detección (LOD) es una medida de la ASe de una prueba. El LOD es la cantidad estimada de analito en una matriz determinada que produciría un resultado positivo por lo menos durante parte del tiempo. Habitualmente, el LOD se estima añadiendo analito a la matriz problema. La elección del/los analito/s (por ejemplo, especies o cepas) forma parte de la definición de la ASe y debe especificarse adecuadamente. Pueden diseñarse estas pruebas en función de la estimación precisa y exacta del nivel de probabilidad (por ejemplo, 50% o 100%), pero en determinadas circunstancias puede ser aceptable una estimación conservadora del LOD (por ejemplo, del 100%). Por ejemplo, en una titulación en la que se empleen diluciones decimales, todas las réplicas a todas las diluciones podrían mostrar una respuesta o bien del 100% o bien del 0%. En este caso existirían dos opciones. La última dilución que presente una respuesta del 100% puede aceptarse como estimación conservadora del límite inferior de detección. Una estimación más exacta puede conseguirse mediante una prueba de segunda fase empleando

intervalos más estrechos en el esquema de dilución, centrándose en la región situada entre el 100% y el 0%. En el Capítulo 2.2.5 *Enfoques estadísticos de la validación* se muestran los métodos de evaluación estadística de los datos del LOD.

1.4. Exactitud analítica de las pruebas o procedimientos complementarios

Algunos métodos o procedimientos analíticos pueden ser calificados para su uso como herramientas analíticas en el laboratorio de diagnóstico. Estas suelen ser pruebas o procedimientos complementarios secundarios que se aplican a un analito que se ha detectado en una prueba primaria. La finalidad de tales instrumentos de análisis es caracterizar en mayor grado el analito detectado en la prueba primaria. Algunos ejemplos de pruebas complementarias abarcan desde la neutralización vírica hasta la tipificación o la secuenciación molecular de un virus aislado y MALDI-TOF-MS (espectrometría de masas por tiempo de vuelo con ionización por desorción láser asistida por matriz) para bacterias (Ricchi *et al.*, 2016).

Estas pruebas complementarias deben validarse en cuanto a las características de rendimiento analítico (Apartados A.2 a B.1.3, arriba), pero difieren de las pruebas diagnósticas en que no requieren validación relativa a las características de rendimiento diagnóstico (Apartados B.2 a B.4, abajo) si sus resultados no se utilizan para establecer un diagnóstico final respecto a la finalidad definida. La exactitud analítica de estas herramientas se puede definir por comparación con un reactivo estándar de referencia, o por las características inherentes a la propia herramienta (como la titulación a punto final). En estos ejemplos, el analito en cuestión se caracteriza en mayor medida cuantitativa o cualitativamente mediante la herramienta analítica.

2. Fase 2 – Rendimiento diagnóstico de una prueba

Las muestras de animales utilizadas para evaluar la DSe y la DSp proceden de cuatro fuentes principales: (1) bancos de referencia con muestras que se sabe si están o no infectadas, (2) muestras de brote o programas de vigilancia cuyos animales de procedencia no se sabe si están infectados, aunque sí se sabe si la población está o no infectada (3) no se sabe si los animales o la población están o no infectados, y (4) estudios de desafío experimental (véanse las ventajas y limitaciones de cada método en la Tabla 3; en la Figura, los grupos 2 y 3 se combinan para abreviar). La DSe (proporción de muestras de animales de referencia que se sabe que están infectados y que dan positivo en una prueba) y de la DSp (la proporción de muestras de animales de referencia que se sabe que no están infectados y que dan negativo en una prueba) son los indicadores principales del rendimiento establecidos durante la validación de una prueba (véanse los Capítulos 2.2.1, 2.2.2. y 2.2.3). Estas estimaciones son la base del cálculo de otros parámetros a partir de los cuales se realizan inferencias sobre los resultados de la prueba (por ejemplo, los valores predictivos y los cocientes de probabilidades de los resultados positivos y negativos de la prueba). Por consiguiente, es muy importante que las estimaciones sobre la sensibilidad y la especificidad diagnósticas sean tan exactas como sea posible. Lo ideal es que deriven del análisis de un conjunto de muestras procedentes de animales, cuyos antecedentes y estado en cuanto a la enfermedad/infección en cuestión se conozcan y sean relevantes para el país o región en los cuales se va a utilizar la prueba. Una estimación del área bajo la curva característica operativa del receptor (ROC) es una herramienta útil para estimar la DSe y la DSp de una prueba diagnóstica cuantitativa porque evalúa la exactitud global teniendo en cuenta todos los posibles valores de la misma (Greiner *et al.*, 2000; Zweig y Campbell, 1993). Este enfoque se describe en detalle en el Capítulo 2.2.5.

Sensibilidad diagnóstica

Porcentaje de animales que se sabe que están infectados y que dan positivo en la prueba; los animales infectados que dan negativo se consideran falsos negativos.

Especificidad diagnóstica

Porcentaje de animales que se sabe que no están infectados y que dan resultados negativos en la prueba; los animales no infectados que dan resultados positivos se

Las muestras de animales utilizadas para evaluar la DSe y la DSp proceden de cuatro fuentes principales: (1) bancos de referencia con muestras que se sabe si están o no infectadas, (2) muestras de brote o programas de vigilancia cuyos animales de procedencia no se sabe si están infectados, aunque sí se sabe si la población está o no infectada), (3) no se sabe si los animales o la población están o no infectados, y (4) estudios de desafío experimental (véanse las ventajas y limitaciones de cada método en la Tabla 3; en la Figura, los grupos 2 y 3 se combinan para abreviar).

Muestras de referencia: El número designado de muestras que se sepa que son positivas y muestras que se sepa que son negativas dependerá de cuáles sean los valores probables de DSe y de DSp de la prueba candidata y del nivel de confianza deseado para las estimaciones (Tabla 1 y Jacobson, 1998). En la Tabla 1, se indican dos conjuntos de las cantidades teóricas de muestras necesarias, permitiendo un error del 5% o del 2% en las estimaciones de la DSe o la DSp. La comparación de un error del 5% frente a un error del 2% presenta una considerable reducción en el número de muestras necesarias. Para lograr una confianza alta (habitualmente, del 95%) en las estimaciones de DSe o DSp se requieren muchas muestras cuando se desea un margen de error pequeño en la estimación. Por ejemplo, al pasar de un error del 2% a un error del 5% para una DSe o DSp probables del 90% y un 95% de confianza se observa un aumento considerable (864 frente a 138) en el número de muestras necesarias. En el caso de las enfermedades de la lista económicamente más importantes, podría ser preferible un intervalo de confianza del 99%. Las limitaciones logísticas y financieras podrían requerir que se evaluara un tamaño de muestra inferior al exigido estadísticamente, en cuyo caso el intervalo de confianza calculado para la DSe y la DSp indicarán menor confianza diagnóstica en los resultados. El tamaño de la muestra también puede resultar limitado por el hecho de que no se disponga de las poblaciones de referencia ni los estándares de referencia de la OMSA (para más información, véase el Capítulo 2.2.5). Por lo tanto, inicialmente puede ser necesario utilizar menos muestras. Sin embargo, puede ser muy deseable potenciar la confianza y reducir el margen de error permitido en las estimaciones de la DSe y la DSp añadiendo más muestras (de estado equivalente respecto al conjunto original) a medida que se dispone de ellas.

Definición de caso, qué constituye un animal infectado y qué constituye un animal no infectado, respectivamente, por ejemplo, signos clínicos, fiebre, resultados de una prueba de referencia, etc.

Tabla 2. Cantidad teórica de muestras procedentes de animales que se sabe si están o no infectados, necesaria para establecer las estimaciones de sensibilidad (DSe) y especificidad (DSp) diagnósticas en función del valor probable de DSe o DSp y del margen de error y el nivel de confianza deseados

Estimación de DSe o DSp	2% de error permitido en la estimación de la DSe y la DSp			5% de error permitido en la estimación de la DSe y la DSp		
	Confianza			Confianza		
	90%	95%	99%	90%	95%	99%
90%	610	864	1493	98	138	239
92%	466	707	1221	75	113	195
94%	382	542	935	61	87	150
95%	372	456	788	60	73	126
96%	260	369	637	42	59	102
97%	197	279	483	32	45	77
98%	133	188	325	21	30	52
99%	67	95	164	11	15	26

Los siguientes son ejemplos de poblaciones y metodologías de referencia que pueden ayudar a estimar las características de rendimiento de la prueba que está siendo validada.

2.1. Poblaciones de animales de referencia

Teóricamente, la selección de animales de referencia requiere que variables importantes del hospedador de la población estudiada estén representadas en los animales escogidos para ser infectados por el agente en cuestión o expuestos al mismo, o que nunca hayan sido infectados ni expuestos (Tabla 4). Las variables a destacar son, aunque no exclusivamente, la especie, la edad, el sexo, la raza, el estadio de la infección, el historial de vacunación y el historial de enfermedades relevantes en el rebaño (para más información, véase el Capítulo 2.2.6). Tras la detección inicial de una nueva enfermedad, puede no haber muestras de referencia o ser escasas. En estas situaciones, las muestras de los estudios experimentales de desafío pueden ser las únicas disponibles para la validación inicial de los ensayos.

2.1.1. Muestras de referencia negativas

Puede ser difícil localizar muestras que se haya comprobado que son negativas, de animales que no hayan tenido posibilidad de infección ni exposición al agente en cuestión. A menudo es posible obtener estas muestras de países o zonas en los que la enfermedad en cuestión se ha erradicado o nunca ha existido. Estas muestras son útiles siempre que la población objeto de la prueba sea suficientemente similar a la población de la que procede la muestra.

2.1.2. Muestras de referencia positivas

En general es problemático hallar cantidades suficientes de animales de referencia verdaderamente positivos, comprobados mediante el aislamiento del agente patógeno. Puede ser necesario recurrir a muestras de animales que se hayan identificado mediante otras pruebas de exactitud suficientemente alta, como las pruebas validadas de detección de ácido nucleico. La prueba candidata se aplica a estas muestras de referencia y los resultados (positivos y negativos) se clasifican de forma cruzada en una tabla 2x2. Esto se ha denominado “modelo de referencia” porque en él se considera que el estándar de referencia es perfecto. En la Tabla 2 del apartado B.2.5) se muestra un cálculo de la muestra. Heuer y Stevenson (2021) describen situaciones en las que, para la validación de pruebas de diagnóstico, se dispone de una referencia perfecta para animales positivos o bien para negativos, y otra en la que la referencia es perfecta para ambos.

2.2. Muestras de animales que no se sabe si están infectados

En algunas situaciones, es posible que se sepa si una población está o no infectada. Si se sabe que la población no está infectada, se supone que todos los animales de esa población tampoco lo están. En una población infectada, no todos los animales estarán infectados, sobre todo si la enfermedad no es muy contagiosa. Deducir la situación de la población respecto a la infección es más difícil si la infección es subclínica o silente o si los animales están protegidos por vacunación o exposición previa a un agente patógeno. Cuando el denominado estándar de referencia (cuando se sabe que hay infección) es imperfecto, que es lo habitual en cualquier prueba de diagnóstico que se utilice para analizar muestras de campo, las estimaciones de la DSe y de la DS_p para la prueba candidata que se basen en este estándar no serán perfectas. Una forma de superar este problema es llevar a cabo un análisis de clases latentes (LCA) de los resultados conjuntos de dos o más pruebas asumiendo que ninguna prueba es perfecta (p. ej. Johnson *et al.*, 2019).

Los modelos de clases latentes no parten de la suposición de una prueba de referencia perfecta, sino que estiman la exactitud de la prueba candidata y de una prueba de referencia utilizando los resultados analíticos conjuntos (Branscum *et al.*, 2005; Enøe *et al.*, 2000; Georgiadis *et al.*, 2003; Hui y Walter, 1980). Si se utiliza un marco de trabajo bayesiano para un LCA, se pueden incorporar al análisis conocimientos previos sobre la DSe y la DS_p de la prueba de referencia y de la prueba candidata. En un LCA bayesiano, también se puede especificar si la población de origen está o no infectada, y la inclusión de muestras de animales de poblaciones que se sepa que no están infectadas puede mejorar la capacidad de estos modelos para estimar la DS_p y otros parámetros del modelo (por ejemplo, la DSe o la prevalencia) con mayor precisión que si sólo se utilizan datos de dos o más poblaciones infectadas. Normalmente, en un LCA deben cumplirse tres supuestos clave: (1) cuando se utilizan datos de múltiples poblaciones, la prevalencia de cada población debe ser diferente; (2) la DSe y la DS_p de la prueba son constantes en todas las poblaciones analizadas; y (3) las pruebas son condicionalmente independientes. Sin embargo, para los casos de dependencia condicional entre las pruebas índice y de referencia (por ejemplo, cuando ambas miden un analito similar), se han desarrollado modelos de clases latentes (LCM, por las siglas en inglés de *latent class models*) con diferente estructura de dependencia para modelar la dependencia condicional entre las pruebas

Dado que los modelos de clase latente bayesianos son complejos y requieren ceñirse a suposiciones críticas, debe solicitarse la ayuda de expertos en estadística para orientar el análisis y describir el muestreo de la/s población/es de destino y las características de otras pruebas incluidas en el análisis, y para que sea correcta la elección del modelo y los métodos de estimación, esta debe basarse en bibliografía revisada por expertos (más información en el Capítulo 2.2.5 y Cheung *et al.*, 2021). El cálculo del tamaño muestral para un LCA requiere de la ayuda de un estadístico o epidemiólogo con experiencia en estas técnicas).

Tabla 3. Poblaciones de origen de las muestras utilizadas para la validación de una prueba

Bancos de muestras de referencia (p. ej., de laboratorios de ref.)	Poblaciones de campo con animales que no se sabe si están infectados	Ensayos experimentales de desafío
<p style="text-align: center;">Ventajas</p> <ul style="list-style-type: none"> • Se sabe si están o no infectadas • Análisis estadístico simple <p style="text-align: center;">Limitaciones</p> <ul style="list-style-type: none"> • A menudo no existen en los casos de enfermedades minoritarias o nuevas • Pueden no representar a animales que se analicen en el futuro • Es posible que no se conozca la situación clínica de los animales de los que se han tomado, los resultados obtenidos en otras pruebas o la información demográfica sobre las mismas 	<p style="text-align: center;">Ventajas</p> <ul style="list-style-type: none"> • Representan a los animales que se analizarán en el futuro • Se utilizan modelos de clase latente (LCM) para el análisis de datos: se puede tener en cuenta la situación de la población respecto a la infección y covariables • Permiten evaluar la exactitud de pruebas múltiples <p style="text-align: center;">Limitaciones</p> <ul style="list-style-type: none"> • El LCM exige pericia y formación, y ceñirse a los supuestos del modelo 	<p style="text-align: center;">Ventajas</p> <ul style="list-style-type: none"> • Permiten determinar si hay infección o no según los resultados de pruebas de desafío realizadas en el grupo • Permiten evaluar la “ventana diagnóstica” de una prueba <p style="text-align: center;">Limitaciones</p> <ul style="list-style-type: none"> • Las muestras de los ensayos de desafío pueden no reflejar las muestras obtenidas sobre el terreno. • A menudo es un resultado secundario en un estudio de la patogénesis. • La vía de exposición, la dosis infecciosa y otros parámetros experimentales pueden influir. • Son más adecuados para infecciones agudas • Consideraciones éticas

2.3. Animales de referencia infectados experimentalmente o vacunados

Las muestras obtenidas secuencialmente a partir de animales infectados experimentalmente o vacunados son útiles para determinar la cinética de las respuestas de anticuerpos o la presencia/ausencia de antígeno o microorganismos en muestras de dichos animales (Tabla 3). Sin embargo, los resultados múltiples adquiridos de forma seriada antes y después de la exposición de animales individuales no son aceptables para establecer estimaciones de la DSe y la DSp porque se infringe el requisito estadístico de las observaciones independientes. Puede aceptarse el muestreo de animales de experimentación realizado en un solo momento (por ejemplo, una muestra tomada aleatoriamente de cada animal). No obstante, es importante tener en cuenta que en el caso de los métodos indirectos de detección del analito, la exposición a microorganismos en condiciones experimentales o la vacunación pueden desencadenar respuestas de anticuerpos que podrían no ser cuantitativa y cualitativamente típicas de la infección natural en la población estudiada (Jacobson, 1998). La cepa del microorganismo, la dosis y la vía de administración a los animales de experimentación son ejemplos de variables que pueden inducir a error cuando se extrapolan las estimaciones de sensibilidad y especificidad diagnósticas a la población estudiada. Si debido a que es prácticamente imposible obtener muestras de referencia adecuadas de animales expuestos de forma natural se hace necesario utilizar muestras de animales de experimentación para los estudios de validación, las mediciones resultantes de DSe y de DSp deben considerarse resultados de “pruebas de demostración del concepto” más que estimaciones teóricas de las verdaderas DSe y DSp.

2.4. Elección de un valor umbral para clasificar los resultados de una prueba

Para obtener estimaciones de la DSe y la DSp de la prueba candidata, que se mide en una escala continua, en primer lugar, los resultados de la prueba deben agruparse en categorías: dos (positivos o negativos) o tres (positivos, intermedios [dudosos] o negativos). Esto se consigue insertando uno o dos puntos de corte (umbrales o límites de decisión) en la escala de resultados de la prueba. Los puntos de corte escogidos deben reflejar la finalidad definida de la prueba y su aplicación, y deben respaldar la DSe y la DSp requeridas para la prueba. Existen opciones y métodos descriptivos para determinar la mejor forma de expresar la DSe y la DSp (Branscum *et al.*, 2005; Georgiadis *et al.*, 2003; Greiner *et al.*, 1995; Greiner *et al.*, 2000; Jacobson, 1998; Zweig y Campbell, 1993 y el Capítulo 2.2.5). Si se produce una

superposición considerable en las distribuciones de los valores de las pruebas realizadas en animales que se sepa que están o no infectados, resulta imposible seleccionar un único punto de corte que permita clasificar adecuadamente a estos animales con relación a su estado de infección. En lugar de un único punto de corte, se pueden seleccionar dos puntos de corte que definan una DSe alta (por ejemplo, que incluya el 99% de los valores de los animales infectados), y una DSp alta (por ejemplo, que incluya el 99% de los valores de los animales no infectados) (Greiner et al., 1995).

La principal dificultad de establecer puntos de corte basados en las características del rendimiento diagnóstico es el no conocer el número necesario de muestras bien caracterizadas (de referencia). Se explican alternativas en el Apartado B.2.6 sobre la aceptación provisional de una prueba durante la recogida de datos para mejorar las estimaciones de la DSe y de la DSp.

2.5. Cálculo de la DSe y de la DSp basado en los resultados del análisis de muestras de referencia

Un método habitual de determinación de las estimaciones de la DSe y la DSp es analizar las muestras de referencia con la prueba nueva, y tabular de forma cruzada los resultados categóricos de la prueba en una Tabla de 2 x 2. En un ejemplo hipotético, supongamos que el técnico que ejecutó la prueba eligió un tamaño de muestra suponiendo que los valores más probables de DSe y de DSp de la prueba nueva son del 97% (DSe) y del 99% (DSp), respectivamente, con una confianza deseada del 95% en ambas estimaciones. La cantidad de error permitido en las estimaciones se fijó en el 2%. La Tabla 1 indica que se necesitan 279 muestras de animales que se sepa que están infectados para evaluar la DSe, y que se necesitan 95 muestras que se sepa que son negativas para establecer la estimación de la DSp. Suponiendo que se analizaran las muestras mediante la prueba nueva, la Tabla 4 es un conjunto hipotético de resultados a partir de los cuales se han obtenido las estimaciones de la DSe y de la DSp.

Tabla 4. Estimaciones de la sensibilidad y la especificidad diagnósticas calculadas a partir de un conjunto hipotético de resultados de muestras analizadas procedentes de poblaciones que se sabe que están infectadas o de poblaciones que se sabe que no están infectadas.

		Cantidad de muestras de referencia necesarias*			
		Que se sabe que son positivas (279)		Que se sabe que son negativas (95)	
Resultados de la prueba	Positivos	270	VP	FP	7
	Negativos	9	FN	VN	88
		Sensibilidad diagnóstica*		Especificidad diagnóstica*	
		VP/(VP + FN)		VN/(VN + FP)	
		96,8% (94,0 – 98,5%)**		92,6% (85,4 – 97,0%)**	

*Basado en la Tabla 1 para una prueba con los siguientes parámetros:

- 1) Antes del análisis, DSe estimada del 97% y DSp estimada del 99%
 - 2) 95% = confianza necesaria en las estimaciones de la DSe y la DSp
 - 3) 2% = Error permitido en las estimaciones de la DSe y la DSp
- VP y FP = Verdadero Positivo y Falso Positivo, respectivamente
VN y FN = Verdadero Negativo y Falso Negativo, respectivamente

**Intervalo de confianza binomial exacto del 95% para los valores calculados de DSe y DSp (véase el Capítulo 2.2.5 para información sobre límites de confianza)

En este ejemplo, las estimaciones de la DSe son las previstas, pero la DSp es muy inferior (92%) al valor previsto del 99%. Como consecuencia, la amplitud del intervalo de confianza para la DSp es muy superior a la esperada. Al volver a examinar la Tabla 1, se observa que son necesarias 707 muestras para lograr un margen de error de $\pm 2\%$ para una DSp del 92%, pero este aumento del tamaño de la muestra podría no ser factible (véase el Capítulo 2.2.5 para más información).

2.6. Reconocimiento provisional de una prueba⁷

En determinadas situaciones no se puede o no se debe pasar a la Fase 2 del Proceso de Validación porque no se dispone de suficientes muestras adecuadas de la población estudiada y es difícil acceder a los animales (por ejemplo, debido a enfermedades infecciosas transfronterizas o a enfermedades de la fauna salvaje).

La experiencia ha mostrado que el principal obstáculo para pasar a la Fase 2 del Proceso de Validación es la cantidad de muestras específicas necesarias para calcular la DSe y la DSp. La fórmula se conoce bien y existen tablas para determinar el número de muestras necesarias para estimar distintos niveles de DSe y DSp, en función del margen de error y del nivel de confianza deseados en las estimaciones (Tabla 2 y Jacobson, 1998). En la fórmula se da por supuesto que está contemplada la infinidad de factores relacionados con el hospedador/microorganismo que pueden influir en el resultado de la prueba. Dado que este supuesto puede ser cuestionable, los tamaños de muestra estimados en el mejor de los casos son muy pequeños. En el caso de una enfermedad que no sea endémica ni esté diseminada, inicialmente podría ser imposible obtener el número de muestras requerido, pero con el tiempo, la acumulación de datos adicionales permitirá el ajuste del punto de corte (umbral) o, en el caso de que no se precise ajuste, una mejora de la confianza en las estimaciones.

El reconocimiento provisional define una prueba en cuya Fase 1 se han evaluado parámetros básicos (ASe, ASp y repetibilidad) y, además, se ha realizado una estimación preliminar de la DSp y la DSe basada

en un conjunto pequeño de muestras bien caracterizadas que contengan el analito de interés y una estimación preliminar de la reproducibilidad. Esto supone haber terminado parte de la Fase 2. Pueden llevarse a cabo estimaciones preliminares de la reproducibilidad de la prueba candidata empleando el grupo escogido de muestras bien caracterizadas para mejorar el estado de aceptación provisional de la prueba. A continuación, el método analítico candidato se duplica en laboratorio de al menos dos centros distintos, y el conjunto de muestras se evalúa empleando la muestra candidata en cada uno de estos laboratorios, utilizando el mismo protocolo, los mismos reactivos que se han descrito en el protocolo y un equipo comparable. Es una versión reducida del estudio de reproducibilidad de la Fase 3 de la validación de la prueba. Siguiendo este procedimiento de reconocimiento provisional, el protocolo de la prueba no debe variar.

Reconocimiento provisional

Prueba con ASe y ASp aceptables, repetibilidad y DSe y DSp preliminares prometedoras basadas en un pequeño panel escogido de muestras bien caracterizadas que contengan el analito objetivo y una estimación preliminar de la reproducibilidad. Los resultados pueden reconocerse de forma provisional hasta que se obtengan muestras de tamaño suficiente tanto para la DSe como para la DSp (véase la Tabla 2) y se realice un estudio de la reproducibilidad más amplio para confirmar la idoneidad general para las finalidades definidas.

El reconocimiento provisional de una prueba por parte de autoridades estatales o nacionales significa que no se han evaluado las características de rendimiento diagnóstico de la prueba. Así, el laboratorio debe desarrollar y seguir un protocolo para añadir y evaluar muestras, a medida que se disponga de ellas, con el fin de cumplir este requisito. Teóricamente, este proceso debe limitarse a un periodo de tiempo concreto en el cual esta recopilación de muestras tendría por objetivo cumplir las Fases 2 y 3 del proceso de validación, y se reservaría a situaciones determinadas (emergencias, especies menores, situaciones en las que no se disponga de ninguna otra prueba, etc.).

3. Fase 3 – Estimaciones de la reproducibilidad y del aumento de la repetibilidad

3.1. Reproducibilidad

La reproducibilidad es la capacidad de un método analítico de que sus resultados sean coherentes, según las estimaciones de la precisión, al aplicarlo a alícuotas de las mismas muestras analizadas en distintos laboratorios, preferiblemente situados en regiones o países distintos utilizando exactamente la misma prueba (protocolo, reactivos y controles). Para evaluar la reproducibilidad de una prueba, cada uno de al menos tres laboratorios debe analizar el mismo conjunto de un mínimo de 20 muestras, con

⁷ El reconocimiento provisional de una prueba no implica la aceptación por parte de la OMSA. No obstante, reconoce una decisión informada de las autoridades a nivel local, estatal, nacional o internacional de que aprueba de forma condicional una prueba validada parcialmente.

idénticas alícuotas para cada laboratorio (véase el Capítulo 2.2.6). Este ejercicio también genera datos preliminares sobre efectos no aleatorios atribuibles a la utilización de la prueba en otros laboratorios. Además, las estimaciones de la repetibilidad intra-laboratorio aumentan por las réplicas que se utilizan en los estudios de reproducibilidad. Pueden estimarse las mediciones de la precisión de los datos tanto de reproducibilidad como de repetibilidad (en el Capítulo 2.2.4 se explica en mayor detalle este tema y su aplicación). Johnson y Cabuang (2021) y Waugh y Clark (2021) presentan factores que afectan a la reproducibilidad de las pruebas entre laboratorios y ejemplos prácticos de pruebas de competencia y pruebas comparativas entre laboratorios. En Ludi et al. (2021) se presenta un estudio de caso sobre la fiebre aftosa para la elección y el uso de paneles de referencia, y en Watson et al. (2021) se describe el valor de los biobancos virtuales a los efectos de transparencia con respecto a los reactivos y las muestras utilizados durante el desarrollo y la validación de las pruebas.

En el caso de las pruebas a pie de establo (POCT, por las siglas en inglés de *point-of-care test*), la reproducibilidad debe evaluarse en las condiciones en que pretenda utilizarse cada prueba.

3.2. Designación de prueba validada

Al terminar la validación de Fase 3, suponiendo que las fases anteriores se hayan superado total y satisfactoriamente, la prueba puede considerarse “validada para la finalidad inicial definida”. El que se mantenga esta designación depende de si se lleva a cabo un seguimiento continuo del rendimiento de la prueba, como se describe en el apartado 5.1.

4. Fase 4 – Implementación del programa

Cuando una prueba funciona bien, se obtiene información adicional y útil de su rendimiento respecto a las expectativas. Además, la prevalencia (real) del rasgo diagnóstico en la población en estudio es un factor importante que debe tenerse en cuenta como se describe abajo.

4.1. Idoneidad

Aunque este capítulo trata la validación y la idoneidad para una finalidad definida desde el punto de vista científico, también debe tenerse en cuenta que existen otros factores que podrían influir en la utilidad de una prueba para la aplicación prevista. Estos factores son no solo la idoneidad diagnóstica de la prueba, sino también su aceptabilidad por parte de las comunidades científica y reguladora, la aceptabilidad para el cliente y la viabilidad dados los recursos de laboratorio de los que se disponga. En algunas enfermedades, tal vez se disponga de varias pruebas que puedan utilizarse junto con programas de control y vigilancia de la enfermedad, en cuyo caso, la utilidad de la prueba tal vez tenga que valorarse evaluando incrementos graduales en los valores de la DSe, la DS_p y predictivos de las dichas pruebas combinadas.

Cuando resulta imposible cumplir los requisitos de funcionamiento de una prueba, dicha prueba no resultará idónea para la finalidad definida. Estos requisitos pueden ser costes de funcionamiento, disponibilidad de equipo, nivel de sofisticación técnica y de capacidad de interpretación, disponibilidad de kits/reactivos, periodo de validez, requisitos de transporte, seguridad, bioseguridad, rendimiento de la muestra, tiempo necesario para la entrega de resultados, aspectos del control de calidad y de la garantía de calidad, y si es viable llevar a cabo la prueba en otros laboratorios. Los kits de las pruebas utilizados sobre el terreno son muy deseables en el sentido de que son muy fáciles de utilizar, pero dado que se usan fuera del entorno controlado de un laboratorio, deben añadirse precauciones para que sigan siendo adecuados para la finalidad en cuestión (Crowther et al., 2006; Halpin et al., 2021). En la “Tabla 1 Propósitos de las pruebas e importancia proporcional de la sensibilidad diagnóstica (DSe), la especificidad diagnóstica (DS_p), el valor predictivo positivo (PV+), el valor predictivo negativo (PV-), el cociente de probabilidades de un resultado positivo (LR+) y el cociente de probabilidades de un resultado negativo (LR-)” figuran ejemplos que respaldan cada una de las seis finalidades.

4.2. Interpretación de los resultados de la prueba

Valores predictivos de los resultados de las pruebas: El valor predictivo positivo (PPV) es la probabilidad de que un animal que haya dado positivo en la prueba realmente sea positivo según el diagnóstico real. El valor predictivo negativo (NPV) es la probabilidad de que un animal que ha dado negativo en la prueba de hecho sea negativo según el diagnóstico real.

Los valores predictivos de los resultados de una prueba son una aplicación del teorema de Bayes y se calculan del siguiente modo:

$$PPV = \frac{P \times DSe}{P \times DSe + (1 - P) \times (1 - DSp)} \quad \text{y} \quad NPV = \frac{(1 - P) \times DSp}{P \times (1 - DSe) + (1 - P) \times DSp}$$

Donde:

PPV = Valor predictivo de un resultado positivo en una prueba

NPV = Valor predictivo de un resultado negativo en una prueba

P = Prevalencia de infección

DSe = Sensibilidad diagnóstica

DSp = Especificidad diagnóstica

Contrariamente a lo que ocurre con la DSe y la DSp, los valores predictivos están influidos por la prevalencia real del estado real de la población en estudio respecto a la infección. En otras palabras, los valores predictivos no son características inherentes a una prueba diagnóstica determinada, sino una función de su DSe y su DSp y de la prevalencia local de infección en una población definida en un momento determinado.

Los valores predictivos resultan de gran importancia para que los veterinarios de campo interpreten los resultados. Por ejemplo, un PPV de 0,9 significa que un animal que ha dado positivo en la prueba tiene un 90% de probabilidad de estar realmente infectado y un 10% de probabilidad de ser un falso positivo.

El valor predictivo de un resultado positivo también es muy importante para los servicios veterinarios responsables de la gestión de los programas de control o erradicación. En cuanto a la inversa del PPV (es decir, 1/PPV), da la información sobre cuánto dinero se gasta en el desvío de verdaderos y falsos positivos por cada animal realmente positivo que se detecta mediante las actividades de vigilancia. Dicho de otra forma, si el PPV es 0,67, significa que dos de cada tres animales positivos son realmente positivos y que el restante es un falso positivo. Dado que, durante la aplicación de un programa de control, la prevalencia de infección está cambiando continuamente, el seguimiento del PPV es una forma de evaluar los costes del programa.

Además, durante la aplicación de un programa de control, normalmente es aconsejable cambiar la sensibilidad de las pruebas utilizadas; en base a la variación de la prevalencia de infección en la población estudiada y al objetivo del programa, el PPV puede utilizarse para aplicar cambios en la DSe y la DSp en función de aspectos económicos. En otras palabras, cuando surge la necesidad de aplicar un cambio en la DSe y la DSp de la prueba, pueden fijarse varios umbrales posibles a lo largo de la curva ROC de validación de la prueba, y pueden utilizarse los valores pertinentes de DSe y DSp para cada punto de corte para evaluar el coste esperable del desvío de cada animal infectado. Si la finalidad es establecer evidencia de la ausencia de enfermedad, el NPV es la medición más importante. El NPV depende totalmente de la DSe.

Aunque los valores predictivos pueden ser una herramienta útil en la interpretación de las pruebas de diagnóstico, dependen en gran medida de la prevalencia de la enfermedad en la población. Los valores predictivos calculados en poblaciones con prevalencia alta, o en el pico de un brote en que la prevalencia de la enfermedad sea alta, no son aplicables a poblaciones con prevalencia baja o en la cola de un brote en que la prevalencia de la enfermedad haya disminuido notablemente.

La razón de verosimilitud (RV) indica la capacidad diagnóstica de un resultado de una prueba y puede utilizarse como ayuda para la interpretación de la prueba. El cociente de probabilidades es una característica inherente a la prueba; depende únicamente de la sensibilidad y de la especificidad diagnósticas combinadas y, por lo tanto, no varía con la prevalencia. Los cocientes de probabilidades son extremadamente potentes, ya que pueden utilizarse para calcular la probabilidad "post-prueba" de enfermedad, teniendo en cuenta el resultado cuantitativo de la prueba y la evaluación de la probabilidad de infección llevada a cabo por el diagnosticador antes de realizar la prueba (Caraguel y Colling, 2021).

El LR se calcula como la relación entre la probabilidad de que un determinado resultado de la prueba se produzca en individuos infectados y la probabilidad de que el mismo resultado se produzca en individuos no infectados. A la inversa, si el LR es inferior a uno, el resultado de la prueba apoya la ausencia de la infección, es decir, es menos probable que este resultado de la prueba se produzca en animales

infectados que en animales no infectados. Un LR igual a “1” significa que el resultado de la prueba no tiene poder diagnóstico (es decir, es tan probable que se produzca en animales infectados como en animales no infectados). Cuanto más se aleje el LR de uno, hacia cero o infinito, mayor será la evidencia proporcionada por el resultado de la prueba. En el contexto clínico, los resultados de las pruebas con un $LR > 10$ o $< 0,1$ se consideran una buena prueba diagnóstica de la presencia o ausencia de la infección, respectivamente.

$$LR^+ = \frac{DSe}{1 - DSp} \quad \text{y} \quad LR^- = \frac{1 - DSe}{DSp}$$

Donde:

LR^+ = Coeficiente de probabilidad de un resultado positivo en la prueba

LR^- = Cociente de probabilidad de un resultado negativo en la prueba

DSe = Sensibilidad diagnóstica

DSp = Especificidad diagnóstica

El LR oscila entre cero e infinito. Si el LR de un resultado dado es mayor que uno, este resultado de la prueba respalda la presencia de infección. Los LR pueden aplicarse en el punto de corte o en diferentes rangos de resultados de la prueba.

4.3. Reconocimiento internacional

Tradicionalmente, la OMSA ha reconocido pruebas a nivel internacional cuando están diseñadas como pruebas prescritas o alternativas para fines comerciales. Esto a menudo se ha basado en la evidencia de su utilidad a nivel nacional, regional o internacional. En el caso de los kits de diagnóstico comerciales que han pasado por el procedimiento de validación y certificación de pruebas de diagnóstico de la OMSA, el paso final es incluir la prueba en el Registro de la OMSA. Las pruebas incluidas en el Registro están certificadas como adecuadas para una finalidad definida si han superado las fases de Validación 1, 2 y 3 seguidas de una revisión de respaldo por parte de un panel de expertos independientes. El Registro tiene por objetivo proporcionar a los posibles usuarios de la prueba una fuente informada e imparcial de información sobre el kit y sus características de rendimiento para una finalidad definida (Gifford *et al.*, 2021). Este Registro puede consultarse en la página web de la OMSA⁸.

4.4. Utilización de la prueba

La prueba definitiva de la utilidad de una prueba es/son su/s aplicación/es con éxito en otros laboratorios y su inclusión en programas nacionales, regionales y/o internacionales de control o vigilancia. Los laboratorios de referencia desempeñan un papel crucial en este proceso (Brown *et al.*, 2021). En el avance natural de las mejoras diagnósticas y/o tecnológicas, pruebas validadas recientemente podrían convertirse en el nuevo método de referencia frente al cual se compararán las demás. Como tales, poco a poco obtendrán el reconocimiento nacional, regional e internacional. Como estándares reconocidos, estas pruebas también se utilizarán para elaborar reactivos de referencia con fines de control de calidad, eficiencia y armonización. Esos reactivos de referencia también pueden convertirse en estándares internacionales.

Debe repetirse una evaluación de la reproducibilidad cuando la prueba se transfiere del laboratorio al campo, tanto para su uso en laboratorios locales como en aplicaciones a pie de explotación. Los cambios predecibles, como extremos de temperatura y niveles de experiencia del técnico, deben evaluarse como fuentes adicionales de variación en los resultados de la prueba que pueden afectar a las estimaciones de la reproducibilidad.

5. Seguimiento del rendimiento de una prueba tras la validación inicial

5.1. Seguimiento de la prueba

Para que una prueba siga considerándose validada, es necesario asegurarse de que la misma conserva las características de rendimiento tal como se definieron durante la validación (véase el apartado 6,

8 [Kits de Diagnóstico - OMSA - Organización Mundial de Sanidad Animal \(woah.org\)](https://www.woah.org/)

abajo), lo cual puede determinarse mediante un programa de garantía de calidad que se caracterizará por un seguimiento cuidadoso del rendimiento diario de la prueba, principalmente con estimaciones de la precisión y la exactitud obtenidas al analizar los controles internos, así como de la tendencia a dar datos atípicos (1.1. Repetibilidad). Puede realizarse un seguimiento del rendimiento gráficamente marcando en diagramas de control los resultados obtenidos con los controles⁹ (Crowther et al., 2006). Deben estudiarse las desviaciones respecto al rendimiento esperado para poder aplicar medidas correctivas si es necesario. Este seguimiento aporta evidencias cruciales de que la prueba conserva su calificación de “validada” durante la fase de implementación de la prueba. La reproducibilidad se evalúa mediante programas de control de calidad externos, como la comprobación de la competencia (Johnson y Cabuang, 2021) (3.1. Reproducibilidad). En caso de que la prueba empiece a dar resultados que no concuerden con los datos de la validación inicial, deberá clasificarse como no adecuada para la finalidad definida. Así, las pruebas validadas deben someterse a una evaluación continua para asegurarse de que siguen siendo adecuadas para la finalidad definida (Waugh y Clark, 2021).

5.2. Modificaciones y mejoras – consideraciones a tener en cuenta para aplicar cambios en la prueba

Con el tiempo, es probable que sea necesario aplicar modificaciones en la prueba para abordar posibles cambios en la finalidad de la prueba, en los analitos estudiados (es decir, modificaciones en la prueba para ajustar el rendimiento diagnóstico) o bien en las características técnicas, con el fin de mejorar la eficiencia o la relación coste-eficacia de la prueba. En caso de que se produzca un cambio en la finalidad de la prueba, será obligatoria una validación revisada desde la Fase 2 en adelante.

Si la prueba va a aplicarse a otra región geográfica y/o población, es recomendable volver a validarla en las nuevas condiciones. Se sabe que los linajes o sublinajes de los agentes infecciosos de los animales varían en función de la región geográfica, lo cual implica tener que revalidar la prueba para cada población en estudio. Esto es especialmente cierto en el caso de los sistemas de detección de ácido nucleico (DAN) y es muy frecuente que se produzcan mutaciones puntuales en muchos agentes infecciosos (sobre todo, en virus de ARN). Las mutaciones, que pueden tener lugar en puntos del cebador o de la sonda, pueden afectar a la eficiencia de la prueba e incluso invalidar las características de rendimiento establecidas. También es aconsejable confirmar periódicamente la secuencia de acceso en las regiones genómicas escogidas de cepas nacionales o regionales de los agentes infecciosos. Esto es especialmente cierto para los puntos del cebador y de la sonda, para garantizar que permanecen estables y que la DSe y la DSp de la prueba no resultan comprometidas. Pueden surgir cuestiones similares con las pruebas basadas en la inmunología, tanto para detección de antígenos como de anticuerpos.

Puede producirse una situación similar con el surgimiento de nuevos subtipos de agentes patógenos existentes. En tales circunstancias, las pruebas DAN existentes tal vez tengan que modificarse.

5.2.1. Modificaciones técnicas y evaluaciones de la comparabilidad

Lo habitual es que las modificaciones técnicas realizadas en una prueba validada, como los cambios de instrumental o de los protocolos de extracción, así como la conversión de una prueba en un sistema semiautomático o totalmente automático utilizando la robótica, no exijan una completa revalidación de la prueba. En lugar de ello, se lleva a cabo una comparación de los métodos para determinar si las modificaciones relativamente pequeñas de la prueba afectan a las características de rendimiento documentadas con anterioridad. La comparabilidad puede establecerse ejecutando el procedimiento modificado y el original en paralelo, con el mismo conjunto de muestras en ambos, y realizando varias ejecuciones. El conjunto de muestras escogido para esta comparación debe representar el intervalo completo de funcionamiento de ambas pruebas. Si se determina que los resultados del procedimiento modificado y del método original validado son comparables en un experimento basado en un criterio pre-especificado, la prueba modificada sigue considerándose validada para la finalidad definida. Véase el Capítulo 2.2.8 por una descripción de experimentos que son adecuados para comprobar la

9 *Gráfico de control:* Es una representación gráfica de datos obtenidos de la medición repetida de una o varias muestras control analizadas en distintas ejecuciones de la prueba a lo largo del tiempo.

comparabilidad, el Capítulo 2.2.6. sobre conjuntos de muestras de referencia, Bowden y Wang, 2021 y Reising et al. 2021.

5.2.2. Modificaciones biológicas y evaluaciones de la comparabilidad

Puede haber situaciones en que sea necesario y/o esté justificado realizar cambios en algunas de las sustancias biológicas utilizadas en la prueba. Estos cambios pueden realizarse en la muestra en sí (por ejemplo, una tejido distinto o incluso otra especie animal), en los reactivos (por ejemplo, la sustitución de un antígeno recombinante por un antígeno derivado de cultivo celular, o de un antígeno conjugado a anticuerpo por otro de especificidad inmunológica similar en un ELISA). La dificultad de cualquier modificación radica en el hecho de que debe determinarse si el cambio requiere una revalidación completa de la prueba tanto a nivel de laboratorio como de campo. Como mínimo, toda modificación requiere se evalúen los “requisitos analíticos” apropiados de la Fase 1. La decisión más difícil es la relativa al “rendimiento diagnóstico” de la Fase 2. Para facilitar dicha decisión, en primer lugar, la prueba original (de referencia) debe compararse con la prueba modificada (candidata) en un ensayo controlado en el que se empleará un conjunto definido de muestras diagnósticas positivas y negativas. En el Capítulo 2.2.8 y en Reising et al. (2021) se ofrece una descripción de la evaluación de la comparabilidad. Si la evaluación de la comparabilidad no sugiere un cambio en el rendimiento diagnóstico, la prueba modificada se puede pasar a la fase de uso sistemático. Si, por el contrario, se observan diferencias en la DSp y la DSe, para adoptar la prueba modificada se requerirá otra Fase 2 o una validación sobre el terreno.

Se precisa un **estudio de la comparabilidad** cuando se ha producido un cambio en el protocolo analítico de una prueba validada, con el fin de asegurar que el rendimiento de la prueba sea comparable.

5.2.3. Reposición de reactivos agotados

Cuando un reactivo, como una muestra control o un estándar de trabajo, está a punto de agotarse, es fundamental preparar y analizar repetidamente un reactivo de repuesto antes de que aquellos se agoten. La futura muestra control debe incluirse en múltiples ejecuciones de la prueba paralelamente con el control original para establecer su relación de proporcionalidad. Siempre que sea posible, es importante cambiar cada vez solamente un reactivo para evitar agravar el problema al tener que evaluar más de una variable.

5.3. Mejora de la confianza en los criterios de validación

Dado que muchas de las variables del hospedador influyen en el rendimiento diagnóstico de las pruebas, es muy deseable aumentar con el tiempo el número de muestras de referencia o de muestras adecuadas para un análisis de clases latentes. El diseño de la obtención, recogida, transporte y entorno en el que se analicen nuevas muestras deben ser los mismos que los aplicados al estudio de validación inicial. Al aumentar el número de muestras mejora la precisión de las estimaciones globales de la DSe y la DSp, y puede permitir cálculos de las estimaciones de la DSe mediante factores como la edad, el estadio de la enfermedad y la carga de microorganismos. La experiencia práctica demuestra que es difícil actualizar periódicamente los expedientes de validación con nuevos datos. La participación en rondas de pruebas de competencia utilizando paneles pertinentes con las cepas más recientes puede ayudar a demostrar la exactitud y precisión continuas de un ensayo.

5.3.1. Gestión de los datos

El almacenamiento de datos a largo plazo, la revisión de los datos de validación y la verificación continua son componentes importantes para aumentar la confianza en que los ensayos siguen rindiendo con una exactitud diagnóstica aceptable. Para alcanzar este objetivo, se necesitan sistemas de gestión de la información de laboratorio (LIMS, por las siglas en inglés de *Laboratory Information Management Systems*) que integren los datos de validación y diagnóstico para proporcionar actualizaciones periódicas del rendimiento de los ensayos. Como mínimo, los sistemas de gestión de datos deben facilitar (1) el almacenamiento de los datos de validación en ubicaciones centrales, (2) la revisión y/o el intercambio de datos cuando, por ejemplo, los ensayos se desplieguen en laboratorios externos (Fase 4 del proceso de validación de ensayos de la OMSA (véase la Figura 1), (3) acumulación e integración continuas de resultados de diagnóstico para actualizar las características de diagnóstico de un ensayo

(Fase 3, véase la Figura 1), y (4) almacenamiento de los datos de los controles de calidad interno (IQC) y externo (EQC) para su uso en la revisión de la sensibilidad del ensayo (Fase 1, véase la Figura 1), especialmente cuando se producen cambios en los reactivos o en el equipo.

El acceso a los LIMS y su utilización para la validación y la verificación continua de pruebas siguen siendo limitados, ya que requieren una inversión importante y conocimientos informáticos para su mantenimiento.

6. Verificación de pruebas validadas

Si un laboratorio está planteándose utilizar un kit comercial validado o una prueba candidata en base a la bibliografía publicada sobre datos de validación, será preciso algún tipo de verificación para determinar si dicha prueba cumple con lo afirmado por el fabricante o el autor respecto a los criterios de validación de la Fase 1 y en el contexto de la aplicación pretendida. Ello puede requerir una verificación reducida tanto de la A_{Sp} como de la A_{Se} empleando materiales de referencia de los que se disponga, tanto si se han adquirido fuera como dentro de la población en estudio. Una vez el laboratorio está seguro de que la prueba, desde el punto de vista analítico, está rindiendo como se ha descrito, para que pueda pasar a utilizarla de forma sistemática deberá plantearse el pasar antes por una Fase 2 reducida en el contexto de la aplicación pretendida y de la población en estudio (Kirkland y Newberry, 2021).

Se precisa un **estudio de verificación** cuando una prueba validada se utiliza en otro laboratorio, con el fin de garantizar que los resultados del estudio de validación original pueden verificarse y que el rendimiento global de la prueba es comprable.

7. Nuevas tecnologías

El uso de la secuenciación de alto rendimiento (HTS, por las siglas en inglés de *high-throughput sequencing*) y las oportunidades para su aplicación al diagnóstico están aumentando rápidamente; los principales objetivos son la secuenciación imparcial para el descubrimiento de agentes patógenos y la secuenciación dirigida para la detección y posterior caracterización de los mismos. Si el ensayo se utiliza para detectar microorganismos no identificados previamente, por ejemplo, durante la investigación de un brote, el objetivo principal es el diagnóstico. Si el ensayo se utiliza para caracterizar mejor un patógeno previamente identificado o para seguir la epidemiología molecular del patógeno durante un brote, entonces su finalidad general puede describirse como una prueba auxiliar. En ausencia de un analito diana conocido, no es posible seguir una vía de validación tradicional. Para evaluar el rendimiento relativo de varios ensayos HTS se utilizan métricas de calidad de seguimiento como la profundidad de la cobertura, la uniformidad de la cobertura, el sesgo de GC, las puntuaciones de calidad de la llamada de base, la disminución de la intensidad de la señal o la longitud de la lectura, la calidad del mapeo y la inclusión de controles internos (Halpin *et al.*, 2021; van Borm *et al.*, 2016).

Otro aspecto complejo en la validación de pruebas de diagnóstico es el desarrollo de ensayos multiplexados, como los ensayos basados en microesferas y la RT-PCR/PCR en tiempo real multiplexada, en los que se identifica más de una diana. Es importante, aunque difícil, validar la precisión de la prueba respecto a cada una de esas dianas a diferentes concentraciones y distribuciones.

8. Conclusiones

Una adecuada validación y verificación adecuadas de las características de rendimiento de las pruebas de diagnóstico de enfermedades infecciosas es fundamental para garantizar que los ensayos se apliquen e interpreten de una manera científicamente sólida y defendible (Colling y Gardner, 2021). Desde que, en 1996, se adoptó por primera vez el Procedimiento de Desarrollo y Validación de Pruebas de Diagnóstico de la OMSA (Figura 1), se ha considerado la norma reconocida internacionalmente para la validación de pruebas de diagnóstico de enfermedades infecciosas de los animales.

Aunque la OMSA esboza un enfoque integral para la validación de pruebas, la experiencia ha demostrado que los desarrolladores de ensayos siguen enfrentándose a una serie de desafíos críticos para cumplir con muchas de las Fases de dicho procedimiento. Al identificar estos obstáculos, se han detectado oportunidades de mejora en las normas de validación, los métodos y la ampliación de las actividades relacionados con las pruebas de diagnóstico, todo lo cual se resume en la Tabla 5 (Reid *et al.*, 2022).

Tabla 5. Resumen de retos y oportunidades en la validación de las pruebas de diagnóstico

Retos	Oportunidades
<ul style="list-style-type: none"> ¿Cuál es el objetivo de la prueba? 	<ul style="list-style-type: none"> Definir claramente la finalidad o finalidades y los parámetros asociados pertinentes; por ejemplo, una prueba de cribado requiere una sensibilidad diagnóstica alta y una prueba de confirmación requiere una especificidad diagnóstica alta.
<ul style="list-style-type: none"> Definir el alcance y las limitaciones de la prueba 	<ul style="list-style-type: none"> ¿Qué se puede esperar de la prueba (alcance) y qué no (limitaciones)?
<ul style="list-style-type: none"> Definición de caso: ¿qué constituye un animal infectado y qué constituye un animal no infectado? 	<ul style="list-style-type: none"> Por ejemplo: positivo en la(s) prueba(s) de referencia, lesiones características, infección experimental, muestra tomada de un individuo de una población históricamente negativa, etc. si no se sabe si está infectado. Puede ser aplicable el modelo bayesiano de clases latentes (BLCM)
<ul style="list-style-type: none"> ¿Tiene la prueba de referencia una precisión insuficiente o probablemente inferior a la nueva prueba evaluada? 	<ul style="list-style-type: none"> Utilizar el BLCM para estimar la sensibilidad y la especificidad del diagnóstico, los cocientes de probabilidades y otros parámetros pertinentes, como la prevalencia.
<ul style="list-style-type: none"> Especies y muestras 	<ul style="list-style-type: none"> Definir la especie y la muestra para las que se validará la prueba, por ejemplo, pollo doméstico, hisopos nasofaríngeos.
<ul style="list-style-type: none"> Origen y población diana 	<ul style="list-style-type: none"> ¿Es la población de origen (donde se validó la prueba) similar a las poblaciones objetivo (donde se aplicará la prueba)?
<ul style="list-style-type: none"> Diseño, análisis, interpretación y presentación de informes de los estudios de validación/verificación (falta de datos de validación originales) 	<ul style="list-style-type: none"> Capítulos 1.1.6 y 2.2.1-2.2.8 del <i>Manual Terrestre</i>, proceso de certificación y registro de la OMSA, la declaración STARD (<i>Standards for Reporting of Diagnostic Accuracy Studies</i>), STARD para la paratuberculosis, STARD para los animales acuáticos, BLCM, ISO/IEC 17025:2017, plantillas de validación de organizaciones nacionales e internacionales y talleres proporcionados por los Laboratorios de Referencia y Centros Colaboradores de la OMSA.
<ul style="list-style-type: none"> Falta de muestras (enfermedades nuevas y emergentes, minoritarias, subclínicas, de la fauna salvaje, etc.) 	<ul style="list-style-type: none"> Paneles de referencia de muestras bien descritas, si las hay, y muestras para la comparación entre laboratorios (colaboración en red, "Vetlab"), reconocimiento provisional

La minuciosidad con la que deben validarse los ensayos puede parecer una tarea desalentadora para los laboratorios de diagnóstico y los equipos de investigación. Sin embargo, existen recursos para ayudar a planificar y orientar los estudios de validación y verificación, como los capítulos pertinentes del *Manual Terrestre* de la OMSA, el reciente número especial de la *Revista Científica y Técnica* de la OMSA sobre Validación de Pruebas de Diagnóstico (Colling y Gardner, 2021) y los documentos de orientación publicados por organismos de acreditación nacionales y regionales. Los líderes internacionales en el campo de la validación de pruebas diagnósticas, incluida la OMSA y sus Laboratorios de Referencia y Centros Colaboradores asociados, así como los organismos reguladores nacionales, tienen un papel importante que desempeñar en el desarrollo continuo de las normas y los sistemas necesarios para garantizar que los desarrolladores de ensayos dispongan de los recursos necesarios y los incentivos para cumplir con su responsabilidad de desarrollar e informar sobre estudios de validación bien diseñados y transparentes. Los usuarios finales de las pruebas de diagnóstico también deben recibir apoyo para asumir la responsabilidad de comprender y verificar el rendimiento de los ensayos en sus propios laboratorios, y comunicar claramente la incertidumbre asociada a los resultados de las pruebas de diagnóstico a las partes interesadas.

BIBLIOGRAFÍA

BATH C., SCOTT M., SHARMA P.M., GURUNG R.B., PHUENTSHOK Y., PEFANIS S., COLLING A., SINGANALLUR N., FIRESTONE S.M., UNGVANIJBAN S., RATTHANOPHART J., ALLEN J., RAWLIN G., FEGAN M. & RODONI B. (2020). Further development of a reverse-transcription loop-mediated isothermal amplification (RT-LAMP) assay for the detection of Foot-and-Mouth Disease Virus and validation in the field with use of an internal positive control. *Transbound. Emerg. Dis.*, **67**, 2494–2506. <http://dx.doi.org/10.1111/tbed.13589>.

BOWDEN T.R., CROWTHER J.R. & WANG J. (2021). Review of critical factors affecting analytical characteristics of serological and molecular assays. *Rev. Sci. Tech. Off. Int. Epiz.*, **40**, 53–73. doi:10.20506/rst.40.1.3208.

- BOSSUYT P.M., REITSMA J.B., BRUNS D.E., GATSONIS C.A., GLASZIOU P.P., IRWIG L., LIJMER J.G. MOHER D., RENNIE D., DE VET H.C.W., KRESSEL H.Y., RIFAI N., GOLUB R.M., ALTMAN D.G., HOOFT L., KOREVAAR D.A., COHEN J.F. & STARD (Standards for Reporting Diagnostic Accuracy (STARD)) GROUP (2015). STARD. An updated list of essential items for reporting diagnostic accuracy studies. *BJM*, 351:h5527. doi: 10.1136/bmj.h5527.
- BRANSCUM A.J, GARDNER I.A. JOHNSON. W.O. (2005). Estimation of diagnostic-test sensitivity and specificity through Bayesian modelling. *Prev. Vet. Med.*, **68**, 145–163.
- BROEMELING L.D. (2011a). Bayesian Methods for Medical Test Accuracy. *Diagnostics*, **1**, 1–35; <https://doi.org/10.3390/diagnostics1010001>.
- BROEMELING L.D. (2011b). *Diagnostics*, **1**, 53–76; <https://doi.org/10.3390/diagnostics1010053>.
- BROWN I., SLOMKA M.J., CASSAR C.A., McELHINNEY L.M. & BROUWER A. (2021). The role of national and international veterinary laboratories. *Rev. Sci. Tech. Off. Int. Epiz.*, **40**, 159–172. doi:10.20506/rst.40.1.3215.
- CARAGUEL C.G.B. & COLLING A. (2021). Diagnostic likelihood ratio – the next generation of diagnostic test accuracy measurement. *Rev. Sci. Tech. Off. Int. Epiz.*, **40**, 299–309. doi:10.20506/rst.40.1.3226.
- CHEUNG A., DUFOUR S., JONES G., KOSTOULAS P., STEVENSON M.A., SINGANALLUR N.B. & FIRESTONE S.M. (2021). Bayesian latent class analysis when the reference test is imperfect. *Rev. Sci. Tech. Off. Int. Epiz.*, **40**, 271–286. doi:10.20506/rst.40.1.3224.
- COLLING A. & GARDNER I.A. (eds). (2021). Diagnostic test validation science: a key element for effective detection and control of infectious animal diseases (*Rev. Sci. Tech. Off. Int. Epiz.*, **40**). The Special Issue is available at: <https://doc.woah.org/dyn/portal/index.xhtml?page=alo&aloid=41245&req=21&cid=1c1f3a2e-2399-408c-946f-fb8d0c089a57>.
- COLLING A. & GARDNER I.A. (2021). Conclusions: Validation of tests of WOAHL-listed diseases as fit-for-purpose in a world of evolving diagnostic technologies and pathogens. *Rev. Sci. Tech. Off. Int. Epiz.*, **40**, 311–317. doi: 10.20506/rst.40.1.3227
- COLLING A., LUNT R., BERGFELD J., HALPIN K., McNABB L, JUZVA S., NEWBERRY K., MORRISY C., HLAING LOH M., CARLILE G., WAUGH C., WRIGHT L., WATSON J., EAGLES D., LOOMES C., WARNER S., DIALLO I., KIRKLAND P., BRODER C., ZUELKE K., MCCULLOUGH S. & DANIELS P. (2018). A network approach for provisional assay recognition of a Hendra virus antibody ELISA: test validation with low sample numbers from infected horses. *J. Vet. Diag. Invest.*, **30**, 362–369. <https://doi.org/10.1177/1040638718760102>.
- CROWTHER J.R., UNGER H. & VILJOEN G.J. (2006). Aspects of kit validation for tests used for the diagnosis and surveillance of livestock diseases: producer and end-user responsibilities. *Rev. sci. tech. Off. int. Epiz.*, **25**, 913–935. doi:10.20506/rst.25.3.1706.
- CULLINANE A. & GARVEY M. (2021). A review of diagnostic tests recommended by the World Organisation for Animal Health Manual of Diagnostic Tests and Vaccines for Terrestrial Animals. *Rev. Sci. Tech. Off. Int. Epiz.*, **40**, 75–89. <https://doi.org/10.20506/rst.40.1.3209>.
- DEJAEGER B. & VANDER HEYDEN Y. (2006). Robustness tests. *LCGC Europe*, **19**, online at <http://www.lcgceurope.com/lcgceurope/content/printContentPopup.jsp?id=357956>
- DRUCE J., GARCIA K., TRAN T., PAPADAKIS G. & BIRCH C. (2012). Evaluation of swabs, transport media, and specimen transport conditions for optimal detection of viruses by PCR. *J. Clin. Microbiol.*, **50**, 1064–1065. doi:10.1128/JCM.06551-11.
- ENOE C., GEORGIADIS M.P. & JOHNSON W.O. (2000). Estimating the sensitivity and specificity of diagnostic tests and disease prevalence when the true disease state is unknown. *Prev. Vet. Med.*, **45**, 61–81.
- FINDLAY J.W.A. & DILLARD R.F. (2007). Appropriate calibration curve fitting in ligand binding assays. *AAPS J.*, **9** (2): E260–E267. (Also on-line as *AAPS Journal* [2007]; **9** [2], Article 29 <https://www.springer.com/journal/12248>).
- FOORD A., BOYD V., WHITE J., WILLIAMS D., COLLING A. & HEINE H. (2015). Flavivirus detection and differentiation by a microsphere array assay. *J. Virol. Methods*, **203**, 65–72.

- GARCIA M.E., BLANCO J.L., CABALLERO J. & GARGALLO-VIOLA D. (2002). Anticoagulants interfere with PCR used to diagnose invasive aspergillosis. *J. Clin. Microbiol.*, **40**, 1567–1568. doi:10.1128/jcm.40.4.1567-1568.2002.
- GARDNER I.A., COLLING A., CARAGUEL C.G., CROWTHER J.R., JONES G., FIRESTONE S.M. & HEUER C. (2021). Introduction: validation of tests for OIE-listed diseases as fit-for-purpose in a world of evolving diagnostic technologies. *Rev. Sci. Tech. Off. Int. Epiz.*, **40**, 19–28. doi:10.20506/rst.40.1.3207.
- GARDNER I.A., COLLING A. & GREINER M. (2019). Design, statistical analysis and reporting standards for test accuracy studies for infectious diseases in animals: Progress, challenges and recommendations. *Prev. Vet. Med.*, **162**, 46–55. doi:10.1016/j.prevetmed.2018.10.023.
- GARDNER I.A. & GREINER M. (2006). Receiver-Operating Characteristic Curves and Likelihood Ratios: Improvements over Traditional Methods for the Evaluation and Application of Veterinary Clinical Pathology Tests. *Vet. Clin. Pathol.* **35**, 8–17.
- GARDNER I.A., NIELSEN S.S., WHITTINGTON R.J., COLLINS M.T., BAKKER D., HARRIS B., SREEVATSAN S., LOMBARD J.E., SWEENEY R., SMITH D.R., GAVALCHIN J. & EDA S. (2011). Consensus-based reporting standards for diagnostic test accuracy studies for paratuberculosis in ruminants. *Prev. Vet. Med.*, **101**, 18–34. PMID: 21601933.
- GARDNER I.A., STRYHN H., LIND P. & COLLINS M.T. (2000). Conditional dependence between tests affects the diagnosis and surveillance of animal diseases. *Prev. Vet. Med.*, **45**, 107–122. doi: 10.1016/s0167-5877(00)00119-7.
- GEORGIADIS M., JOHNSON, W., GARDNER I. & SINGH R. (2003). Correlation-adjusted estimation of sensitivity and specificity of two diagnostic tests. *Appl. Statist.*, **52** (Part 1), 63–76.
- GIBB A.P. & WONG S. (1998). Inhibition of PCR by agar from bacteriological transport media. *J. Clin. Microbiol.*, **36**, 275–276. doi:10.1128/JCM.36.1.275-276.1998.
- GIFFORD G., SZABO M., HIBBARD R., MATEO D., COLLING A., GARDNER I. & ERLACHER-VINDEL E. (2021). Validation, certification and registration of certified tests and regulatory control of veterinary diagnostic test kits. *Rev. Sci. Tech. Off. Int. Epiz.*, **40**, 173–188. doi:10.20506/rst.40.1.3216.
- GREINER M. & GARDNER I.A. (2000). Epidemiologic Issues in the Validation of Veterinary Diagnostic Tests. *Prev. Vet. Med.*, **45**, 3–22.
- GREINER M., PFEIFFER D. & SMITH R.D. (2000). Principles and practical application of the receiver operating characteristic (ROC) analysis for diagnostic tests. *Vet. Prev. Med.*, **45**, 23–41.
- GREINER M., SOHR D. & GÖBEL P. (1995). A modified ROC analysis for the selection of cut-off values and the definition of intermediate results of serodiagnostic tests. *J. Immunol. Methods*, **185**, 123–132.
- HALPIN K., TRIBOLET L., HOBBS E.C. & SINGANALLUR N.B. (2021). Perspectives and challenges in validating new diagnostic technologies. *Rev. Sci. Tech. Off. Int. Epiz.*, **40**, 145–157. doi:10.20506/rst.40.1.3214.
- HOBBS E., COLLING A., GURUNG R. & ALLEN J. (2020). The potential of diagnostic point-of-care tests (POCTs) for infectious and zoonotic animal diseases in developing countries: technical, regulatory and sociocultural considerations. *Transbound Emerg Dis.*, **68**, 1835–1849.
- HEUER C. & STEVENSON M.A. (2021). Diagnostic test validation studies when there is a perfect reference standard. *Rev. Sci. Tech. Off. Int. Epiz.*, **40**, 261–270. doi:10.20506/rst.40.1.3223.
- HUI S.L. & WALTER S.D. (1980). Estimating the error rates of diagnostic tests. *Biometrics*, **36**, 167–171.
- INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (2019). ISO/TS 22583:2019(en), Guidance for supervisors and operators of point-of-care testing (POCT) devices. <https://www.iso.org/obp/ui/#iso:std:iso:ts:22583:ed-1:v1:en>
- JACOBSON R.H. (1998). Validation of serological assays for diagnosis of infectious diseases. *Rev. sci. tech. Off. int. Epiz.*, **17**, 469–486.

JIA B., COLLING A., STALLKNECHT D.E., BLEHERT D., BINGHAM J., CROSSLEY B., EAGLES D. & GARDNER I.A. (2020). Validation of laboratory tests for infectious diseases in wild mammals: review and recommendations. *J. Vet. Diagn. Invest.*, **32**, 776–792. doi:10.1177/1040638720920346.

JOHNSON P. & CABUANG L. (2021). Proficiency testing and ring trials. *Rev. Sci. Tech. Off. Int. Epiz.*, **40**, 189–203. <https://doi.org/10.20506/rst.40.1.3217>

JOHNSON W.O., JONES G. & GARDNER I. (2019). Gold standards are out and Bayes is in: implementing the cure for imperfect reference tests in diagnostic accuracy studies. *Prev. Vet. Med.*, **167**, 113–127. doi:10.1016/j.prevetmed.2019.01.010.

KIRKLAND P.D. & NEWBERRY K.M. (2021). Your assay has changed – is it still ‘fit for purpose’? What evaluation is required? *Rev. Sci. Tech. Off. Int. Epiz.*, **40**, 205–215. <https://doi.org/10.20506/rst.40.1.3218>.

KOSTOULAS P., GARDNER I.A., ELSCHNER M.C., DENWOOD M.J., MELETIS E. & NIELSEN S.S. (2021). Examples of proper reporting for evaluation (Stage 2 validation) of diagnostic tests for diseases listed by the World Organisation for Animal Health, *Sci. Tech. Off. Int. Epiz.*, **40**, 287–298. doi:10.20506/rst.40.1.3225.

KOSTOULAS P., NIELSEN S.S., BRANSCUM A.J., JOHNSON W.O., DENDUKURI N., DHAND N.K., TOFT N. & GARDNER I.A. (2017). STARD-BLCM: Standards for the Reporting of Diagnostic accuracy studies that use Bayesian Latent Class Models. *Prev. Vet. Med.*, **138**, 37–47. PMID: 28237234.

LUDI A.B., MIOULET V., BAKKALI KASSIMI L., LEFEBVRE D.J., DE CLERCQ K., CHITSUNGO E., NWANKPA N., VOSLOO W., PATON D.J. & KING D.P. (2021). Selection and use of reference panels: a case study highlighting current gaps in the materials available for foot and mouth disease. *Rev. Sci. Tech. Off. Int. Epiz.*, **40**, 239–251. doi:10.20506/rst.40.1.3221

MAYO C.E., WEYER C.T., CARPENTER M.J., REED K.J., RODGERS C.P., LOVETT K.M., GUTHRIE A.J., MULLENS B.A., BARKER C.M., REISEN W.K. & MACLACHLAN N.J. (2021). Diagnostic applications of molecular and serological assays for bluetongue and African horse sickness. *Rev. Sci. Tech. Off. Int. Epiz.*, **40**, 91–104. doi:10.20506/rst.40.1.3210.

MICHEL A.L., VAN HEERDEN H., PRASSE D., RUTTEN V., DAHOUK S. AL & CROSSLEY B.M. (2021). Pathogen detection and disease diagnosis in wildlife: challenges and opportunities. *Rev. Sci. Tech. Off. Int. Epiz.*, **40**, 105–118. doi:10.20506/rst.40.1.3211.

MIYACHI H., MASUKAWA A., OHSHIMA T., FUSEGAWA H., HIROSE T., IMPRAIM C. & ANDO Y. (1998). Monitoring of inhibitors of enzymatic amplification in polymerase chain reaction and evaluation of efficacy of RNA extraction for the detection of hepatitis C virus using the internal control. *Clin. Chem. Lab. Med.*, **36**, 571–575. doi:10.1515/CCLM.1998.098.

NEWBERRY K.M. & COLLING A. (2021). Quality standards and guidelines for test validation for infectious diseases in veterinary laboratories? *Rev. Sci. Tech. Off. Int. Epiz.*, **40**, 227–237. doi:10.20506/rst.40.1.3220.

PÉREZ L.J., LANKA S., DESHAMBO V.J., FREDRICKSON R.L. & MADDOX C.V. (2020). A validated multiplex Real Time PCR assay for the diagnosis of infectious *Leptospira* spp.: a novel assay for the detection and differentiation from both pathogenic groups I and II. *Front. Microbiol.*, **11**, 457. <https://doi.org/10.3389/fmicb.2020.00457>.

REID T., SINGANALLUR N. B., NEWBERRY K., WAUGH C., BOWDEN T. & COLLING A. (2021). Validation of diagnostic tests for infectious diseases: challenges and opportunities. International Symposium on Sustainable Animal Production and Health Current Status and Way Forward. 28 June–2 July 2021, Joint FAO/IAEA Centre. Accepted for publication in symposium proceedings 7 April 2022.

REISING M.M., TONG C., HARRIS B., TOOHEY-KURTH K.L., CROSSLEY B., MULROONEY D., TALLMADGE R.L., SCHUMANN K.R., LOCK, A.B. & LOIACONO C.M. (2021). A review of guidelines for evaluating a minor modification to a validated assay. *Rev. Sci. Tech. Off. Int. Epiz.*, **40**, 217–226. doi:10.20506/rst.40.1.3219.

RICCHI M., MAZARELLI A., PISCINI A., DI CARO, A, CANNAS A., LEO S., RUSSO S. 1 ARRIGONI N. (2016). Exploring MALDI-TOF MS approach for a rapid identification of *Mycobacterium avium* ssp. paratuberculosis field isolates. *J. Appl. Microbiol.*, **122**, 568–577.

SAAH A.J. & HOOVER D.R. (1997). ‘Sensitivity’ and ‘Specificity’ Reconsidered: The Meaning of These Terms in Analytical and Diagnostic Settings. *Ann. Internal Med.*, **126**, 91–94.

SCHRADER C., SCHIELKE A., ELLERBROEK L. & JOHNE R. (2012). PCR inhibitors – occurrence, properties and removal. *J. Appl. Microbiol.*, **113**, 1014–1026. doi:10.1111/j.1365-2672.2012.05384.x.

STEVENSON M., HALPIN K. & HEUER C. (2021). Detection of emerging infectious zoonotic diseases. *Rev. Sci. Tech. Off. Int. Epiz.*, **40**, 119–130. doi:10.20506/rst.40.1.3212.

TRIBOLET, L., KERR, E. COWLED, C., BEAN, A.G., STEWART, C.R., DEARNLEY, M. AND FARR, R. (2020). MicroRNA Biomarkers for infectious diseases: from basic research to biosensing. *Front. Microbiol.*, **11**, 1197. doi: 10.3389/fmicb.2020.01197.

VAN BORM S., WANG J., GRANBERG F. & COLLING A. (2016). Next-generation sequencing workflows in veterinary infection biology: towards validation and quality assurance. *Rev. Sci. Tech. Off. Int. Epiz.*, **35**, 67–81. doi:10.20506/rst.35.1.2418.

VESSMAN J., STEFAN R., VAN STADEN J., DANZER K., LINDNER W., BURNS D., FAJGELJ A. & MULLER H. (2001). Selectivity in analytical chemistry. *Pure Appl. Chem.*, **73**, 1381–1386.

WATSON J.W., CLARK G.A. & WILLIAMS D.T. (2021). The value of virtual biobanks for transparency purposes with respect to reagents and samples used during test development and validation. *Rev. Sci. Tech. Off. Int. Epiz.*, **40**, 253–259. doi:10.20506/rst.40.1.3222.

WAUGH C. & CLARK G. (2021). Factors affecting test reproducibility among laboratories. *Rev. Sci. Tech. Off. Int. Epiz.*, **40**, 131–143. doi:10.20506/rst.40.1.3213.

WILSON I.G. (1997). Inhibition and facilitation of nucleic acid amplification. *Appl. Environ. Microbiol.*, **63**, 3741–3751. doi:10.1128/AEM.63.10.3741-3751.1997.

YAN L., TOOHEY-KURTH K.L., CROSSLEY B.M., BAI J., GLASER A.L., TALLMADGE R.L. & GOODMAN L.B. (2020). Inhibition monitoring in veterinary molecular testing. *J. Vet. Diagn. Invest.*, **32**, 758–766. doi:10.1177/1040638719889315

YOKOTA M., TATSUMI N., NATHALANG O., YAMADA T. & TSUDA I. (1999). Effects of heparin on polymerase chain reaction for blood white cells. *J. Clin. Lab. Anal.*, **13**, 133–140. doi:10.1002/(sici)1098-2825(1999)13:3<133::aid-jcla8>3.0.co;2-0.

ZWEIG M.H. & CAMPBELL G. (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin. Chem.*, **39**, 561–577.

<https://www.iaea.org/services/networks/vetlab>

<https://www.iaea.org/services/zodiac>

<https://www.woah.org/en/what-we-offer/veterinary-products/diagnostic-kits/the-register-of-diagnostic-kits/>

<https://www.awe.gov.au/agriculture-land/animal/health/laboratories/tests/test-development>

*
* *

NB: Existe un Centro colaborador para la Ciencia de Validación de las Pruebas de Diagnóstico en la región Asia-Pacífico (puede consultarse en la página web de la OMSA: <https://www.woah.org/es/que-ofrecemos/red-de-expertos/centros-colaboradores/>).

Para más información sobre la validación por favor contactar con el Centro Colaborador de la OMSA.

NB: ADOPTADO POR PRIMERA VEZ EN 1996 BAJO EL TÍTULO *PRINCIPIOS DE VALIDACIÓN DE LAS PRUEBAS DE DIAGNÓSTICO DE LAS ENFERMEDADES INFECCIOSAS*; ÚLTIMAS ACTUALIZACIONES ADOPTADAS EN 2023.