

CAPÍTULO 1.1.2.

PRINCIPIOS Y MÉTODOS DE VALIDACIÓN DE LAS PRUEBAS DE DIAGNÓSTICO DE LAS ENFERMEDADES INFECCIOSAS

1. Introducción

La validación es un proceso que determina la idoneidad de una prueba, que se ha desarrollado, optimizado y estandarizado adecuadamente para un fin concreto. 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 etapas de la validación (véase la Figura, 1 abajo), incluida la caracterización del rendimiento, puede designarse como “validada para el objetivo inicial deseado(s)”¹. Pero para que una prueba siga estando validada, es necesario realizar un cuidadoso seguimiento del rendimiento de la misma a diario, a menudo controlando el comportamiento de los controles internos a lo largo del tiempo. Esto garantiza que la prueba, según la validación original, mantenga siempre sus características de rendimiento. En el caso de que empiece a dar resultados que no concuerden con los datos de la validación original, la prueba podría no ajustarse al fin deseado. Así, una prueba validada se evalúa continuamente para garantizar que mantiene su idoneidad para el fin deseado, mediante una evaluación de los resultados de los controles internos cada vez que se lleva a cabo.

Los términos ensayo, método analítico y prueba son sinónimos a efectos de este capítulo, y por tanto pueden utilizarse indistintamente.

Los términos “**válido**” (adjetivo) o “**validez**” (sustantivo) indican si las estimaciones de las características de rendimiento de la prueba están sesgadas respecto a los verdaderos valores del parámetro. Estos términos son aplicables con independencia de que la medición sea cuantitativa o cualitativa.

Las pruebas realizadas en individuos o en poblaciones tienen varios propósitos, tales como: documentar la ausencia de una determinada enfermedad en un país o región, evitar su propagación a través del comercio, 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 uno o varios fines deseados optimizando sus características de rendimiento para cada uno, 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.

El siempre cambiante repertorio de nuevos y específicos reactivos de diagnóstico, junto con la existencia de muchos nuevos formatos y protocolos analíticos ha dado lugar a debates sobre cómo validar adecuadamente estas pruebas. Para orientar a los usuarios de pruebas en cuanto a la validación de pruebas más complejas, como las de detección de ácido nucleico, ya no basta con ofrecer simples ejemplos de pruebas serológicas, como el enzimoanálisis. Con el fin de aportar coherencia al proceso de validación para todos los tipos de pruebas, este capítulo se centra en los criterios que deben cumplirse durante la realización y la validación de pruebas de cualquier tipo. La inclusión de la ejecución de la prueba como parte del proceso de validación de la misma puede parecer contra-intuitiva, pero en realidad, tres de los criterios de validación 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 la misma, y ambos contienen criterios de validación que deben cumplirse. Este capítulo también aporta orientación sobre la evaluación de cada criterio indicando las mejores prácticas científicas, que se especifican en los anexos. Dichas prácticas se adaptan a cada uno de los distintos tipos de prueba (como la detección de ácido nucleico, anticuerpos o antígenos).

1 La **validación** no necesariamente implica que el rendimiento de la prueba cumpla algún valor mínimo o que la prueba tenga un rendimiento comparable a cualquier otra prueba parecida, a no ser que se haya tenido en cuenta específicamente en el diseño del estudio de evaluación de la prueba.

2. Métodos directos e indirectos que requieren validación

El diagnóstico de enfermedades infecciosas se lleva a cabo mediante la detección directa y/o indirecta de agentes infecciosos. Los métodos directos permiten la detección de las partículas de los agentes y/o sus componentes, como ácidos nucleicos, proteínas estructurales o no estructurales, enzimas, etc. Los métodos indirectos ponen de manifiesto respuestas inmunitarias humorales o celulares mediante exposición a agentes infecciosos o a sus componentes. Los métodos indirectos utilizados con más frecuencia en la detección de agentes infecciosos son las pruebas serológicas, tales como la neutralización vírica clásica, el enzimoanálisis (ELISA), la inhibición de la hemaglutinación, la fijación del complemento y métodos más novedosos, como los biosensores, la bioluminometría, la polarización de la fluorescencia o la quimioluminiscencia.

Los métodos más frecuentes de detección directa son el aislamiento o el cultivo *in vitro* de microorganismos viables, la microscopía electrónica, la inmunofluorescencia, la inmunohistoquímica, el ELISA para antígenos, la inmunoelectrotransferencia y los sistemas de detección de ácido nucleico (DAN). Los sistemas DAN incluyen la hibridación de ácido nucleico (NAH), los macro- y micromatrices y las diversas técnicas de amplificación del ácido nucleico, como la reacción en cadena de la polimerasa (PCR), o los métodos de amplificación isotérmica, como la amplificación basada en la secuencia del ácido nucleico (NASBA), y la amplificación isotérmica invasora o mediada por bucle (LAMP). Las pruebas de DAN son cada vez más corrientes y en muchos casos sustituyen el aislamiento del virus y el cultivo de bacterias, en concreto para la detección de agentes cuyo cultivo es difícil o imposible. Las herramientas de DAN también se utilizan como un medio secundario para la identificación muy específica de cepas, grupos o linajes de microorganismos después del aislamiento o el cultivo de virus, bacterias y parásitos. El diagnóstico molecular, como la PCR, no requiere: a) la presencia de organismos replicantes, b) una infraestructura costosa de aislamiento vírico, c) hasta varias semanas para lograr un diagnóstico, ni d) conocimientos especiales, de los que a menudo no se dispone en muchos laboratorios; todas son ventajas prácticas. Estos métodos se han vuelto relativamente baratos, seguros y fáciles de usar (Ballagi-Pordány y Belák, 1996; Belák, 2005; 2007; Belák y Thorén, 2001; Burns *et al.*, 2005; Bustin, 2005; Huggett *et al.*, 2005; Lauerman, 2004; Louie *et al.*, 2000). Diferentes métodos de PCR en tiempo real, robots de extracción de ácidos nucleicos y estaciones de trabajo automatizadas para la DAN y la detección de anticuerpos, antígenos y agentes han dado lugar a un gran repertorio de pruebas de alto rendimiento, robustas y muy rápidas y fiables. Aunque los sistemas DAN a menudo tienen la ventaja de una mayor sensibilidad diagnóstica y especificidad analítica que los métodos de recuperación del agente o los ELISA de captura de antígeno, esta ventaja suele conllevar una mayor dificultad de validación de dichas pruebas.

3. Consideraciones preliminares en el desarrollo y validación de pruebas

La primera consideración es definir el objetivo de la prueba, ya que esto guiará todos los pasos subsiguientes del proceso de validación. Si tenemos en cuenta las variables que pueden afectar a la realización de una prueba, podremos apreciar más claramente los criterios que deben tenerse en cuenta para su validación. Estas variables pueden agruparse en tres 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 el analito en cuestión.

Criterio de validación de la prueba: un rasgo característico de una prueba; un factor, medida o estándar decisivos en los que pueda basarse una opinión o decisión.

La matriz (suero, heces, tejido, etc.) en la que puede residir el analito buscado puede contener inhibidores endógenos o exógenos que impidan el funcionamiento de pruebas dependientes de enzimas, como la PCR o el ELISA. Otros factores que también influyen en la concentración y en la composición de un analito (principalmente anticuerpos) de 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 afectar al analito en cuestión.

Los factores que interfieren en el rendimiento analítico de la prueba son el instrumental, el error técnico, la elección de los reactivos (tanto desde el punto de vista químico como biológico) y su calibración, los límites de exactitud y aceptación de los controles de la prueba, los recipientes y formatos de reacción, la calidad del agua, el pH y la ionicidad de los tampones y los diluyentes, las temperaturas de incubación y sus duraciones, y el error introducido por la detección de analitos estrechamente relacionados. También es importante que los reactivos biológicos estén libres de agentes extraños.

Los factores que pueden influir negativamente en el rendimiento diagnóstico de la prueba se asocian principalmente a la utilización de valores de referencia obtenidos en animales infectados/expuestos o que se sabe que no están infectados, para evaluar la sensibilidad (DSe) y la especificidad diagnóstica (DSp) de la prueba. Esto es especialmente difícil porque el grado en que dichos animales representen todas las variables ambientales y del hospedador que tienen lugar en la población estudiada influye de forma decisiva en la fiabilidad de la interpretación de los resultados. Por ejemplo, los especialistas en serología saben que una prueba que se haya validado mediante muestras de suero extraídas de ganado del norte de Europa puede que no dé resultados válidos en poblaciones claramente distintas de ganado bovino de África. El rendimiento diagnóstico de la prueba se complica aún más cuando no se dispone de los valores de referencia para animales infectados, a menudo porque son imposibles de obtener. En esta situación, la DSe y la DSp se pueden estimar, en determinadas circunstancias, mediante modelos de clase latente (Enoe *et al.*, 2000; Greiner y Gardner, 2000; y Apéndice 1.1.2.5).

4. Criterios de desarrollo y validación de las pruebas

El rendimiento de la prueba se ve afectado por muchos factores, que abarcan desde las primeras etapas del desarrollo hasta la etapa final de evaluación de los resultados cuando la prueba se aplica a poblaciones específicas de animales. Una prueba, por lo tanto, no puede ser considerada válida a menos que se hayan comprobado y confirmado o que haya cumplido un conjunto específico de criterios de validación esenciales (véase el recuadro), ya sea cuantitativa o cualitativamente². La falta de cumplimiento de cualquiera de estos criterios es probable que reduzca la confianza en que la prueba cumpla los objetivos deseados. Los primeros cuatro de estos criterios por lo general se tratan durante la fase de desarrollo de la prueba (la Fase de Desarrollo), y los ocho restantes son evaluados durante las tres primeras etapas de la fase de validación de la prueba (la Fase de Validación), como se describe a continuación.

Criterios de validación de las pruebas

1. Aptitud para el fin deseado
2. Optimización
3. Normalización
4. Robustez
5. Repetibilidad
6. Sensibilidad analítica
7. Especificidad analítica
8. Umbrales (puntos de corte)
9. Sensibilidad diagnóstica
10. Especificidad diagnóstica
11. Reproducibilidad
12. Solidez

5. Fase de desarrollo de la prueba

5.1. Definición de los objetivos deseados en una prueba

La *Norma de la OIE sobre Gestión y Requisitos Técnicos para los Laboratorios que Realizan Pruebas para las Enfermedades Infecciosas* (2008) establece que los métodos analíticos y los procedimientos relacionados deben ser adecuados para las aplicaciones diagnósticas específicas a fin de que los resultados de las pruebas tengan utilidad. En otras palabras, la prueba debe ajustarse al fin deseado³. La capacidad del resultado positivo o negativo de una prueba de predecir de forma precisa el estado de infección y/o exposición de un animal o población de animales es el criterio definitivo para la validación de una prueba. Esta capacidad depende de que una prueba, desarrollada mediante una cuidadosa optimización (apartado 5.2.4) y estandarización (apartado 5.2.7) ofrezca, mediante la acumulación de datos de validación, estimaciones de la DSe y la DSp menos sesgadas y más precisas. Estas estimaciones, junto con los datos basados en la evidencia sobre la prevalencia de la infección en la población analizada, son la base para proporcionar un alto grado de confianza en el valor predictivo de los resultados positivos o negativos. Con el fin de asegurar que los resultados de la prueba proporcionan útiles inferencias diagnósticas sobre el estado de infección/exposición de la población de animales, 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 la misma.

Tal como se esboza en la información básica del *Certificado de pruebas de diagnóstico*, en la página web de la OIE (www.oie.int), el primer paso consiste en escoger un tipo de prueba que tenga posibilidades de ser validada para un uso concreto.

2 Para más detalles sobre terminología, consúltense el glosario en los *Estándares y Directrices de la OIE para los Laboratorios Veterinarios: Enfermedades Infecciosas* (Organización Mundial de Sanidad Animal [OIE], 2008).

3 Esta es una interpretación específica de los requisitos establecidos de forma más genérica en la norma internacional de calidad ISO/IEC 17025:2005 para los laboratorios de análisis (ISO/IEC, 2005). En la Norma de la OIE también se establece que, para que un método analítico pueda considerarse adecuado, debe validarse de forma apropiada, y esta validación debe cumplir con los principios indicados en los capítulos sobre validación del *Manual Terrestre* y del *Manual Acuático*.

5.1.1. Aptitud para el fin deseado

Los fines más frecuentes son los siguientes:

- i) Demostrar la ausencia de infección en una población definida (país/zona/compartimento/rebaño) (prevalencia aparente del 0%):
 - a) “Libre” con y/o sin vacunación,
 - b) Restablecimiento de la ausencia después de los brotes
- ii) Certificar la ausencia de infección o presencia del agente causal en animales concretos o productos utilizados para el comercio o el transporte.
- iii) Erradicar de la enfermedad o eliminación de la infección en poblaciones definidas.
- iv) Diagnosticar de manera confirmativa de casos clínicos o sospechosos (incluye la confirmación de un resultado positivo en una prueba de cribado).
- v) Estimar la prevalencia de la infección o exposición para facilitar el análisis de riesgos (encuestas, estatus sanitario del rebaño, medidas para el control de la enfermedad).
- vi) Determinar el estado inmunitario de animales concretos o de poblaciones (después de la vacunación).

Estos fines incluyen muchas aplicaciones más acotadas y específicas de las pruebas (para más detalles véanse los anexos para cada tipo de prueba). Tales aplicaciones específicas y sus fines concretos deben ser claramente definidos en el contexto de una prueba completamente validada.

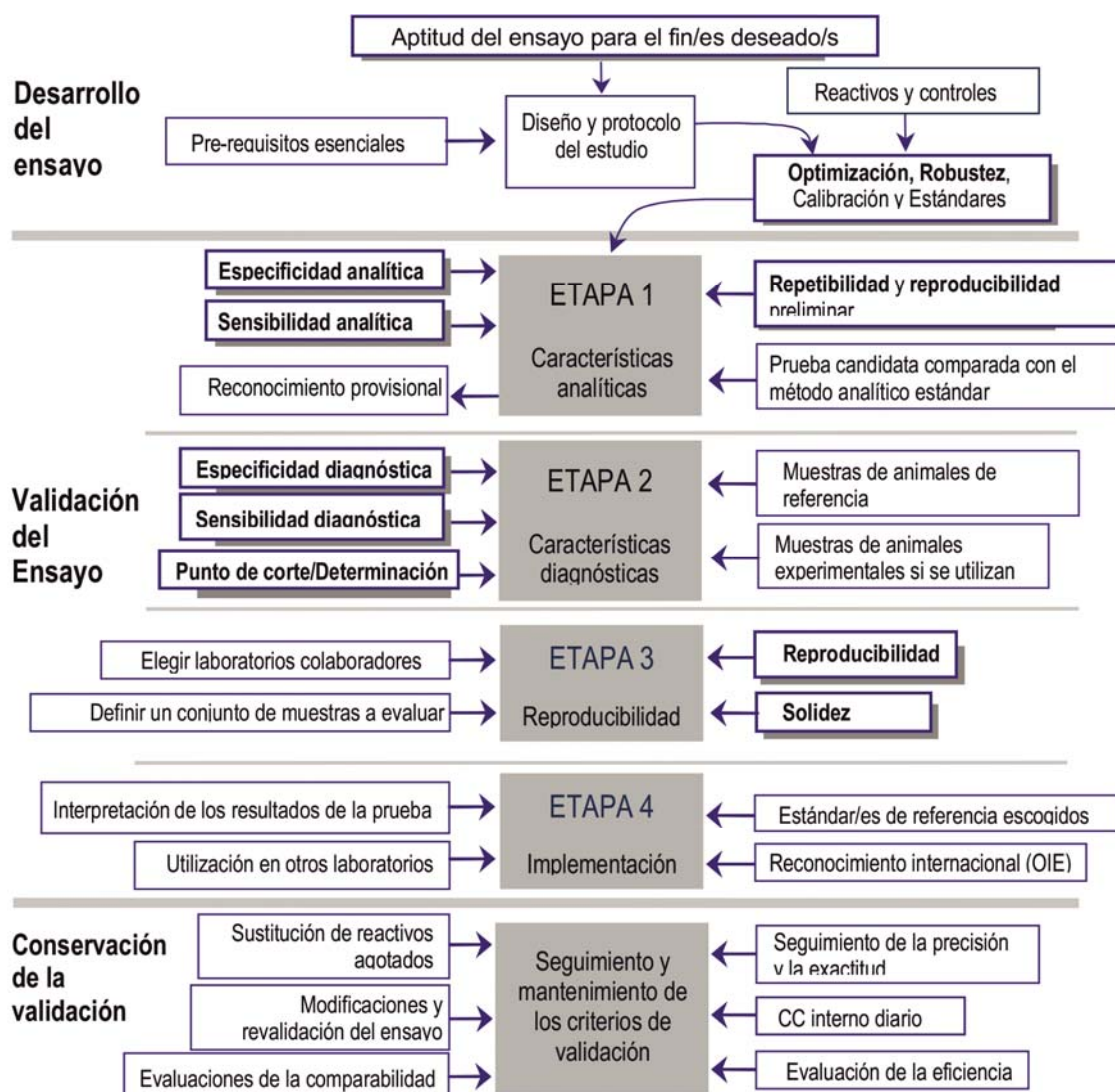


Figura 1. Las fases de desarrollo y validación de la prueba con los criterios de validación de la prueba destacados en negrita en los recuadros sombreados.

5.1.2. Aptitud para la utilización

Si bien el presente capítulo trata de la validación y aptitud para el fin deseado desde un punto de vista científico, debe tenerse en cuenta que existen otros factores prácticos que podrían influir en la utilidad de una prueba para la aplicación deseada. Entre dichos factores se incluyen no sólo la aptitud diagnóstica de la prueba, sino también su aceptación por parte de las comunidades científica y reguladora, su aceptabilidad por parte del cliente, y su aplicabilidad mediante los recursos de laboratorio disponibles. La incapacidad para cumplir con los requisitos operativos de una prueba también puede convertirla en inadecuada para el fin deseado. Entre tales requisitos, se pueden incluir los costes de realización, la disponibilidad de equipo, el nivel de sofisticación técnica y de las habilidades de interpretación, la disponibilidad de kits/reactivos, el periodo de validez, los requisitos de transporte, la inocuidad, la bioprotección, el rendimiento de la muestra, el tiempo de espera para la obtención de resultados, los aspectos relacionados con el control de calidad y la garantía de calidad y si la prueba puede llevarse a cabo en la práctica en otros laboratorios. Los kits de análisis utilizados en el campo son muy deseables desde el punto de vista de facilidad de uso, pero dado que se ejecutan fuera del recinto controlado de un laboratorio, requieren precauciones adicionales para que sigan siendo aptos para el fin deseado (Crowther *et al.*, 2006).

5.2. Desarrollo de la prueba - los estudios experimentales

5.2.1. Pre-requisitos esenciales: factores que influyen en la validación de la prueba

5.2.1.1. Garantía de calidad

Al llevar a cabo pruebas en el laboratorio o al realizar análisis de material clínico, el objetivo es obtener datos de alta calidad. Esto requiere cumplir los requisitos básicos dentro del laboratorio (véanse los capítulos 1.1.3 y 1.1.1 de los *Manuales Terrestre y Acuático*, respectivamente). El establecimiento de los sistemas de garantía de calidad (QA) y de control de calidad (QC) es esencial, y consiste en un conjunto de protocolos de calidad, incluida la utilización de muestras control para la prueba que garanticen que el sistema está funcionando adecuadamente y confirmen la reproducibilidad y calidad de los datos. En muchos laboratorios de todo el mundo ya se han establecido sistemas de QA y QC, y se dispone de personal capacitado y competente.

5.2.1.2. Cómo escoger el equipo

El equipo que no se mantiene y calibra puede constituir un gran impedimento para el logro de una prueba de calidad. Los aparatos (congeladores, bloques de calentamiento, incubadoras, neveras, colorímetros ópticos, termocicladores, lavadores de placas, pipetas, etc.) deben calibrarse según los protocolos de garantía de calidad del laboratorio. Un ejemplo de esta necesidad es la robótica que se utiliza para la automatización de pruebas completas, o partes de las mismas, durante el proceso diagnóstico sistemático. No es suficiente suponer que la extracción robótica de ácido nucleico, por ejemplo, es equivalente a los métodos previos de extracción manual, o que un lavador automatizado de placas de ELISA proporcionará un lavado uniforme de todos los pocillos de las placas. El instrumento debe ser calibrado y el protocolo validado para confirmar la eficiencia de rendimiento y asegurar que no se produzca contaminación cruzada en los sistemas DAN, o que se lleva a cabo un lavado suficiente en todos los pocillos de una placa. (Para más detalles véanse los anexos sobre las mejores prácticas.)

5.2.1.3. Selección e integridad de las muestras

La selección, obtención, preparación y gestión de las muestras son variables críticas en el diseño, desarrollo y validación de una prueba. Otras variables, como el transporte, la cadena de custodia, la trazabilidad de las muestras, y el sistema de gestión de la información del laboratorio también son fuentes importantes de variación/error que cobran especial importancia cuando la prueba se lleva a cabo para análisis sistemáticos. La integridad de los resultados experimentales durante el desarrollo y validación de la prueba tendrá, en el mejor de los casos, la calidad de las muestras utilizadas en las pruebas de diagnóstico experimental o sistemático. Antes de emprender un proceso de validación de una prueba deben preverse los factores que pueden influir negativamente en la calidad de la muestra. Las muestras de referencia utilizadas en el desarrollo y la validación de la prueba deben encontrarse en la misma matriz que se vaya a utilizar en la prueba (por ejemplo, suero, tejido, sangre total) y ser representativas de la especie que se vaya a analizar. Asimismo, los materiales de referencia deben representar adecuadamente el intervalo de concentraciones del analito que tenga que detectar la prueba. Los detalles sobre la obtención, la preparación, el manejo y el transporte adecuados de la muestra se encuentran en la Parte 2. Recomendaciones aplicables a cada enfermedad: Introducción general, y en los Capítulos 2.2.0, 2.3.0 y 2.4.0 de este *Manual Acuático*, en el Capítulo 1.4 del *Código Acuático* y en la Guía de la OIE para la Vigilancia Sanitaria en los animales Acuáticos (2009).

5.2.2. 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 necesarias una buena reflexión y planificación. Se ofrece orientación en los anexos de este capítulo, que cubre las mejores prácticas para el desarrollo y validación de las pruebas destinadas a la detección de distintos analitos (como anticuerpos, antígenos o ácido nucleico).

5.2.2.1. Muestras de analito de referencia

El desarrollo de todas las pruebas depende de las muestras de referencia del analito, que representan el analito y la matriz en la que se encuentra el mismo en la población a la que se destina la prueba. Las muestras de referencia pueden ser sueros, líquidos o tejidos que contengan el analito de interés o una construcción 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.

Las muestras de referencia del analito, que contienen el analito de interés en concentraciones variables, son útiles en el desarrollo y la evaluación de los criterios de validación de la prueba candidata.

5.2.3. Intervalo de funcionamiento de la prueba

Durante el desarrollo de la prueba, se establecen los límites de detección inferior y superior. Para establecer 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 “Optimización”, a continuación). Se lleva a cabo una dilución seriada de esta muestra positiva alta hasta la extinción en una matriz sin analito que tenga la misma constitución que la matriz de las muestras tomadas de 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, OD, Ct, etc.) es función de la concentración (cantidad) de analito. La curva establece el intervalo de funcionamiento de la prueba, que es el intervalo entre la concentración más alta y la más baja (cantidades) de analito en la muestra para el cual se ha demostrado un nivel suficiente de precisión⁴ y exactitud. En la mayoría de las pruebas de diagnóstico, la respuesta es el resultado de la interacción del analito con un anticuerpo u otro reactivo de unión. Estas son conocidos como pruebas de unión de ligando (LBA). La curva de calibración típica de las LBA tiene forma sigmoidea, con un límite inferior (asíntota) cerca de la respuesta mínima (unión inespecífica) y una asíntota superior cerca de la respuesta máxima. Por lo general, los datos de las LBA se transforman para aproximarse a una relación lineal entre la respuesta y la concentración. Esta transformación simplifica la interpolación de los datos mediante análisis de regresión lineal, pero con el inconveniente de que se introduce sesgo. Dado que la linealización es imperfecta porque da lugar a estimaciones comprometidas de la exactitud y la precisión, se han aplicado muchos algoritmos de ajuste de datos a los datos de la curva de calibración experimental de las LBA (Findlay & Dillard, 2007). El modelo de referencia aceptado en la actualidad para la calibración de las LBA es el modelo logístico de 4 parámetros, que por lo general optimiza la exactitud y la precisión en el máximo intervalo de calibración utilizable (Findlay & Dillard, 2007). Tales transformaciones ahora son más prácticas para el usuario general, debido a que en Internet pueden conseguirse muchos programas de estadística fáciles de usar.

La **precisión** es el grado de dispersión entre una serie de mediciones de la misma muestra analizada en las condiciones especificadas (para más detalles, véase la nota 3 al pie)

Intervalo de funcionamiento de una prueba: es un intervalo de concentraciones de analito (cantidades) en el que el método proporciona una precisión y exactitud adecuadas.

La exactitud es la cercanía de un valor obtenido en una prueba al valor esperado (verdadero) de un reactivo de referencia estándar cuya concentración o título se conocen.

5.2.4. Optimización

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, negativo, positivo bajo y positivo alto). En teoría, las muestras deberían representar tanto a los animales que se consideran infectados

4 La precisión se puede evaluar de distintas formas analizando la misma muestra replicada: 1) en una placa o placas en una misma ejecución de la prueba, 2) entre placas analizadas al mismo tiempo en una misma ejecución de la prueba, 3a) entre ejecuciones de la prueba en distintos momentos del mismo día o en distintos días en condiciones similares, 3b) entre ejecuciones de la prueba en distintos días con distintos técnicos, 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 denominan precisión intermedia.

como a los no infectados dentro de la población que finalmente va a ser objeto de la prueba una vez que esta se haya validado. 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, ya que la matriz de las muestras de campo podría ser muy distinta de la matriz de la muestra preparada de este modo. Pero cuando no existe otra alternativa, tal vez la única

La **optimización** es el proceso mediante el cual se evalúan y se ajustan los parámetros físicos, químicos y biológicos más importantes de una prueba para garantizar que las características de rendimiento del mismo se ajusten mejor a la aplicación deseada.

posibilidad sea añadir a la muestra una cantidad conocida del analito 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 y para 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 una prueba de calidad. La evaluación científica y la aplicación de las mejores prácticas científicas indicadas en los anexos de este capítulo son las mejores herramientas para llegar a la optimización de todos los elementos de una prueba. El enfoque descrito proporciona una base firme para el desarrollo de una prueba que cumpla los criterios de “robustez” y “solidez” cuando se utiliza durante largos periodos de tiempo dentro de un laboratorio o cuando se implementa en otros laboratorios. A menudo se han desarrollado ensayos de prototipos 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 optimización se vuelve extremadamente 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 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) a menudo es una fase añadida al final del desarrollo de la prueba, pero debe formar parte de la optimización desde el principio. 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.

En algunos tipos de prueba, el hecho de que el resultado sea correcto depende exclusivamente de dar un determinado paso correcto en el proceso analítico, y requiere una especial atención durante la optimización. Un ejemplo de ello es la extracción de ácido nucleico de la muestra. Para la extracción de ADN o ARN se utilizan métodos químicos tanto comerciales (robótica, columnas de centrifugación, extracciones basadas en imanes, etc.) como estándares. Es crucial determinar el método de extracción más reproducible y eficiente mediante experimentos de optimización. La extracción debe optimizarse para cada tipo de tejido que pueda ser objeto de la prueba. Si el método de extracción se cambia, como mínimo deberá demostrarse una eficiencia de extracción comparable (véase el apartado 6.7, abajo, y el Anexo 1.1.4.6 correspondiente para obtener información adicional sobre cómo establecer la comparabilidad cuando se cambian los reactivos).

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 del mismo. Además, para asegurar la estabilidad tienen que garantizarse la preparación y el almacenamiento adecuados de todos los reactivos biológicos y materiales de referencia. Se ofrecen detalles sobre cómo obtener, preparar, manejar y transportar adecuadamente las muestras en la Parte 2. Recomendaciones aplicables a cada enfermedad: Introducción general, y en los Capítulos 2.2.0,

2.3.0 y 2.4.0 de este *Manual Acuático*, en el Capítulo 1.4 del Código Acuático y en la Guía de la OIE de la Vigilancia Sanitaria en los animales Acuáticos (2009).

Durante la experimentación para optimizar la prueba, debe tomarse nota de los parámetros analíticos cuyo intervalo óptimo de funcionamiento es estrecho, ya que estos son los puntos críticos que pueden afectar a la robustez de una prueba (véase el apartado 5.2.6).

5.2.5. Factores inhibidores en la matriz de la muestra

En general, para la detección de anticuerpos en el suero, las pruebas son bastante resistentes a los factores inhibidores, con la excepción de algunos, como por ejemplo, factores tóxicos en las pruebas de neutralización vírica, o cuando sustancias endógenas de ciertos tipos de muestras inhiben las reacciones enzimáticas en los ELISA. Para la detección de ácidos nucleicos, las matrices de la muestra como la sangre, el suero, los tejidos corporales o las muestras tomadas con hisopos permiten una fácil extracción de los ácidos nucleicos en cuestión, mientras que las heces, tejidos autolisados y las muestras de semen pueden ser más difícil de manejar debido a la presencia de factores que pueden inhibir pruebas posteriores, como la PCR.

5.2.6. Robustez

La robustez es la capacidad de una prueba de no resultar afectada por pequeñas variaciones en las situaciones en que se lleva a cabo el análisis, que podrían tener lugar en el curso de la prueba en un laboratorio determinado. Se evalúa aplicando variaciones intencionadas en los parámetros del método analítico (International Conference on Harmonisation, 2005). 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 presente solidez (presentará una baja reproducibilidad) cuando se transfiera a otros laboratorios (la solidez se trata en el apartado 6.4).

La **robustez** es una medida de la capacidad de una prueba de no resultar afectado por pequeñas, pero intencionadas, variaciones en los parámetros del método, y proporciona una indicación de su fiabilidad durante el uso normal.

Los factores que con mayor probabilidad afectarán a la robustez de la prueba son los cuantitativos (continuos), como el pH y la temperatura; los cualitativos (categóricos), como el lote de los reactivos o la marca de las placas de microtitulación; y los mixtos, como factores relacionados con la matriz acuosa u orgánica (Dejaegher y Vander Heyden, 2006). En las pruebas de unión de ligando (LBA), la falta de robustez no sólo se debe a una concentración/cantidad subóptima del bio-reactivo especificado en el método, sino que también puede deberse a las características intrínsecas del reactivo biológico (como la afinidad por el anticuerpo monoclonal o la avidéz y/o valencia del anticuerpo policlonal). La robustez, por lo tanto, sobre todo de pruebas basadas en LBA, puede verse afectada por errores sistemáticos y/o aleatorios (Thompson *et al.*, 2002).

La comprobación de la robustez se pone de manifiesto método a método. Todos los reactivos críticos se identifican y se someten a una evaluación factorial que compara todas las posibles combinaciones de reactivos. Por ejemplo, en la detección de anticuerpos por ELISA, los factores que se varían pueden ser la concentración de antígeno unido a la fase sólida, la dilución de conjugado y varios sueros problema cuyas concentraciones se sitúen en el intervalo de funcionamiento de la prueba. La respuesta de la prueba ante estos pequeños cambios no debe dar lugar a una variabilidad inaceptable. Por otra parte, la robustez puede comprobarse mediante la aplicación de experimentos de diseño factorial (Dejaegher & Vander Heyden, 2006; Youden y Steiner, 1987)

La robustez se verifica además durante la Etapa 1 de la validación de la prueba. Cuando la prueba optimizada se lleva a cabo por primera vez en condiciones de laboratorio de rutina, esta medida práctica de la robustez se conoce como repetibilidad (véase el apartado 6.2.1) y se supervisa continuamente como parte de los procedimientos de control del proceso durante toda la vida de la prueba (véase el apartado 6.6.1).

5.2.7. Calibración de la prueba frente a reactivos estándar

5.2.7.1. Estándares del analito internacionales y nacionales de referencia

Lo ideal es que los estándares internacionales de referencia, que contienen una concentración conocida del analito, sean los reactivos frente a los cuales todas las pruebas están estandarizadas. Estos estándares los preparan y distribuyen laboratorios internacionales de referencia. Los estándares nacionales de referencia se calibran por comparación con un reactivo estándar internacional siempre que ello es posible, y los prepara y distribuye un laboratorio nacional de referencia. En ausencia de un estándar internacional de referencia, un estándar nacional de referencia se convierte en el estándar de comparación para el prueba candidata. Estos reactivos estándar se caracterizan con gran detalle mediante un exhaustivo análisis, y es preferible optar por métodos de caracterización, preparación y almacenamiento que hayan sido publicados en publicaciones revisadas por pares.

5.2.7.2. Reactivo estándar interno

Un estándar de referencia interno en general tiene la mayor calidad metodológica de la que se dispone en un lugar determinado en una organización dada, y se calibra 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 de los analitos internacionales y nacionales. 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.

5.2.7.3. Reactivo estándar de trabajo

Se calibran uno o más reactivo/s estándar de trabajo, a menudo denominados controles del analito o del proceso, frente a un reactivo 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.

5.2.8. “Normalización” de los resultados de la prueba frente a uno o más estándares de trabajo

Debido a la variación inherente en los resultados brutos de la prueba que se observan con frecuencia entre las ejecuciones de la 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 una 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” [no confundir esto con una transformación de datos para lograr una distribución normal (de Gauss)]. Los valores “normalizados” se pueden expresar de muchas maneras, como en porcentaje de un control positivo (por ejemplo, en un ELISA), o como una concentración o título de un analito derivado de una curva estándar, o como un número de copias genómicas diana también derivadas de una curva estándar de valores Ct (umbral del ciclo) para la PCR en tiempo real. Es una buena práctica para incluir estándares de trabajo (o por lo menos una muestra razonablemente bien caracterizada/s) en todas las realizaciones 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. Los sistemas automáticos de análisis permiten calcular y obtener datos “normalizados” mediante, por ejemplo, una curva estándar, o bien obtener el número de ciclo a partir del cual se está excediendo el umbral de ciclos, como en la PCR en tiempo real. En los Anexos 1.1.4.1 y 1.1.4.3 se ofrece más información.

6. Fase de validación de la prueba

6.1. Definición de prueba validada

La “validación” es un proceso que determina la idoneidad de una prueba que se ha desarrollado, optimizado y estandarizado adecuadamente para un objetivo determinado. 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 etapas de la validación (Figura 1), incluida la caracterización del rendimiento, puede designarse como “validada para el fin original deseado(s)”

Sin embargo, para que conserve el estatus de validada, es necesario asegurar que la prueba mantenga siempre las características de rendimiento definidas durante la validación de la misma (véase el apartado 6.6). Esto se puede determinar en un programa de garantía de la calidad caracterizado por un seguimiento cuidadoso del rendimiento diario de la prueba, principalmente mediante estimaciones de la precisión y la

exactitud de los controles internos, y mediante una comprobación externa programada de la eficiencia. En el caso de que la prueba deje de dar resultados que concuerden con los datos de la validación original, podría no ajustarse al objetivo deseado. Por lo tanto, para garantizar que una prueba validada mantiene su idoneidad para el fin deseado, debe ser continuamente evaluado.

6.2. Etapa 1 – 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. Es posible diseñar experimentos que de manera eficiente proporcionen información sobre fuentes probables de variación en la precisión de la prueba intra e interlaboratorio (véase la nota 4 al pie, en el Apartado 5.2.3), lo cual definirá las características de rendimiento de la prueba. En la etapa 1 deben llevarse a cabo estudios de la repetibilidad, la reproducibilidad y la evaluación de la sensibilidad analítica (límite de detección) con evaluación ciega de los resultados y con una selección aleatoria de las muestras. La elección de los microorganismos, las cepas o los serotipos para evaluar la especificidad analítica deberá reflejar el conocimiento actual y, por lo tanto, informar sobre cuál es el mejor diseño experimental posible para detectar analitos determinados.

6.2.1. Repetibilidad

La repetibilidad se calcula mediante la evaluación de la variación en los resultados de réplicas de un mínimo de tres (y preferiblemente cinco) 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 tres réplicas idénticas de la muestra original que contienen las concentraciones originales de analito y matriz. 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 es 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 (unas 20), llevadas a cabo por dos o más técnicos, preferiblemente en cinco días distintos. La variación en los resultados de las distintas réplicas se pueden expresar como desviaciones estándar, intervalos de confianza, u otras opciones posibles (Véase el Anexo 1.1.4.4 sobre las medidas de incertidumbre para las evaluaciones de la repetibilidad).

La **repetibilidad** es el nivel de concordancia entre los resultados de las réplicas de una muestra dentro y entre las ejecuciones del mismo método analítico en un laboratorio determinado.

6.2.2. Especificidad analítica (ASp)

La especificidad analítica permite a la prueba distinguir entre el analito buscado y otros componentes, al menos de tres formas distintas. Estas se describen como selectividad, exclusividad, e inclusividad de la prueba.

- **La selectividad** es el grado en que un método puede cuantificar con exactitud el analito en cuestión 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); 4) anticuerpos contra la vacunación, que pueden confundirse con anticuerpos contra la infección activa. Estos interferentes pueden causar falsas reducciones o incrementos en 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.
- **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 una prueba confirmativa.
- **La inclusividad** es la capacidad de una prueba de detectar varias cepas o serovares 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.

La **especificidad analítica** es el grado en que la prueba distingue entre el analito en cuestión y otros componentes que pueden ser detectados en la prueba.

Después de eliminar interferentes en la medida de lo posible, el siguiente paso es analizar un conjunto de muestras adecuadas para evaluar la inclusividad, la exclusividad o ambas, dependiendo del propósito de la prueba.

La ASp se aplica tanto a los métodos de detección de analitos directos como a los indirectos. Si se precisa exclusividad, deben obtenerse muestras de campo de animales infectados por microorganismos genéticamente emparentados y no patógenos, pero esto puede resultar difícil o incluso imposible. En tales casos, para los métodos de detección directa pueden utilizarse microorganismos procedentes de cultivos celulares, o bien suero de animales expuestos experimentalmente por vías naturales para los métodos de detección indirecta. El grado de reactividad cruzada que se considera aceptable depende en gran medida del propósito de la prueba y de la prevalencia de los microorganismos/analitos que den reacción cruzada en las muestras de la población objeto de la prueba - lo cual se debe determinar en cada caso (véase el Anexo 1.1.4.5 para más detalles). En el caso de la PCR, es útil realizar estudios de simulación informática como complemento de la evaluación de laboratorio de la ASp; sin embargo, tales estudios no son suficientes por sí mismos para evaluar la ASp.

Un factor propio de la detección de anticuerpos víricos es la posible respuesta humoral de los animales a las proteínas transportadoras que se encuentran en las vacunas -otro tipo de interferente que podría afectar negativamente la selectividad. Si estas proteínas también están presentes en el antígeno de fase sólida en las placas de ELISA, pueden unirse a anticuerpos desarrollados contra proteínas transportadoras de las vacunas y dar resultados falsos positivos (falta de exclusividad en la prueba). Por lo tanto, no se recomienda el uso de preparaciones de vacunas como antígenos en los ELISA. En el Anexo 1.1.4.1 se indican las prácticas específicas para determinar la ASp.

6.2.3 Sensibilidad analítica (ASe)

La sensibilidad analítica es el límite inferior de detección (LOD) de un analito en una prueba. En las pruebas de detección directa, esto se puede expresar como el número de copias del genoma, la dosis infecciosa, las unidades formadoras de colonias, las unidades formadoras de placas, etc. del agente que se pueden detectar y distinguir del resultado de una matriz sin el analito. Por lo general, esto se expresa como un número de copias, unidades de fijación del complemento o unidades formadoras de placa que dan al menos un 50% de resultados positivos entre las réplicas de una muestra de un volumen o peso determinado (véase el Anexo 1.1.4.5 para detalles sobre la determinación del LOD). En las pruebas de detección indirecta, es la menor cantidad de anticuerpos detectados, por lo general, la penúltima dilución de la muestra en la que el analito no es diferenciable de la actividad en una muestra de matriz control.

El límite de detección o sensibilidad analítica es la menor cantidad de analito en una muestra que se puede detectar mediante una prueba de detección directa en al menos el 50% de las réplicas de cada dilución, en una serie de diluciones del analito en la matriz.

Si el propósito es detectar niveles bajos del analito o infecciones subclínicas y es difícil obtener los materiales de referencia adecuados, como por ejemplo muestras de las primeras fases de la infección, podría ser útil determinar la sensibilidad analítica comparativa ejecutando la prueba candidata con un conjunto de muestras y analizando estas muestras también mediante otra prueba independiente. Esto proporcionaría una comparación relativa de la sensibilidad analítica entre las pruebas, pero se debe tener cuidado en la elección de la prueba independiente utilizado en la comparación para asegurarse de que los analitos que se están detectando (si son distintos) tienen el mismo tipo de perfil patógeno en cuanto a tiempo de aparición tras la exposición al agente infeccioso, y a abundancia relativa en las muestras problema escogidas.

Cuando se desarrolla una prueba nueva y más sensible, puede ser necesario analizar muestras seriadas tomadas de animales infectados poco después de que tenga lugar la infección y en adelante a lo largo del desarrollo clínico o hasta que la enfermedad se vuelva fulminante, y analizarlas paralelamente con pruebas previamente utilizadas para demostrar que la sensibilidad es mayor. Esto también proporcionaría una comparación temporal del momento en que se empieza a detectar la infección respecto a la patogénesis de la enfermedad.

6.2.4 Método analítico estándar que se comparará con el método analítico de la prueba candidata

Hay situaciones en las que no es posible o deseable pasar a la Etapa 2 de la Validación porque se dispone de pocas muestras apropiadas de la población estudiada y es difícil acceder a los animales (por ejemplo, en el caso de enfermedades exóticas). Sin embargo, debe analizarse un pequeño pero selecto conjunto de muestras problema altamente caracterizadas que contengan el

Un método analítico estándar es el mejor método reconocido a nivel nacional o internacional o, en su defecto, el mejor método disponible de referencia publicado en una revista de renombre.

intervalo de concentraciones del analito, en paralelo con la prueba candidata y mediante el método estándar, si existe.

6.2.5. Precisión 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. El propósito 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 de un virus aislado o la secuenciación molecular para confirmar una PCR en tiempo real. Los índices de patogenicidad, la inhibición de la hemaglutinación, las determinaciones de resistencia a fármacos, etc., son otros ejemplos en los que se llevan a cabo pruebas o procedimientos complementarios, de forma independiente de la prueba primaria o formando parte de la misma.

Estas pruebas complementarias deben validarse en cuanto a las características de rendimiento analítico (Apartados 5.2 a 6.2.4), pero difieren de las pruebas diagnósticas sistemáticas en que no requieren validación relativa a las características de rendimiento diagnóstico (Apartados 6.3 a 6.5). 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 todos estos ejemplos, el analito en cuestión se caracteriza en mayor medida cuantitativa o cualitativamente mediante la herramienta analítica.

6.2.6 Evaluación preliminar de la reproducibilidad

Las estimaciones preliminares de la reproducibilidad de la prueba candidata pueden ser útiles en este momento en el proceso de validación, donde sólo se dispone de un pequeño conjunto de muestras altamente caracterizadas. Este conjunto de muestras se podría utilizar para una evaluación limitada de la reproducibilidad para potenciar la aceptación provisional de la prueba. A continuación, se duplica el método analítico candidato en los laboratorios de uno o más institutos diferentes, y el conjunto de muestras se evalúa mediante el prueba candidata de cada uno de estos laboratorios, utilizando el mismo protocolo, los mismos reactivos que se especifican en el protocolo, y equipo comparable. Esta es una versión reducida de la Etapa 3 de la validación de la prueba.

6.2.7 Reconocimiento provisional de la prueba⁵

La experiencia ha demostrado que el mayor obstáculo para pasar a la Etapa 2 de la Validación es el número de muestras definido para calcular la DSe y la DSp (véanse los requisitos para la Etapa 2, en rendimiento diagnóstico, abajo). La fórmula es bien conocida y existen tablas para determinar el número de muestras requeridas para estimar los distintos niveles de DSe y DSp, en función de la cantidad de error permitido y del nivel de confianza en las estimaciones (Tabla 1 y Jacobson, 1998). La fórmula asume que se tienen en cuenta todos los numerosos factores del huésped/microorganismo que pueden afectar al resultado de la prueba. Dada que esta suposición puede ser cuestionable, los tamaños estimados de la muestra en el mejor de los casos son muy pequeños. En el caso de las enfermedades que no son endémicas ni generalizadas, inicialmente tal vez sea imposible obtener el número de muestras requerido, pero con el tiempo, la acumulación de datos adicionales permite el ajuste del umbral, y si no es necesario un ajuste, aumenta la confianza en las estimaciones.

El **reconocimiento provisional** se aplica a una prueba en el cual, durante la Etapa 1, se han evaluado los parámetros críticos de referencia: ASe, ASp, repetibilidad y una estimación de la reproducibilidad.

Los precedentes históricos sugerirían que las pruebas en general eran producto de experimentos de laboratorio, con énfasis en la sensibilidad y especificidad analíticas, y que la evaluación de conjuntos de muestras de campo era nominal. Este tipo de validación de referencia para la encefalopatía espongiforme bovina (EEB) es un ejemplo clásico en el que no se disponía de muestras de campo. Sin embargo, durante períodos prolongados de uso en contextos de diagnóstico, estas pruebas han sido objeto de ajustes basados en la evidencia empírica para reducir los resultados falsos positivos y falsos negativos. En algunas de las pruebas rápidas de diagnóstico de la EEB, debían hacerse ajustes en el punto de corte para reducir los resultados falsos positivos evidentes en la implementación inicial. La validación de referencia proporcionó un nivel de confianza en el rendimiento diagnóstico que fue

5 El reconocimiento provisional no implica la certificación por parte de la OIE. Sin embargo, sí implica el reconocimiento de una decisión informada de las autoridades a niveles local, estatal, nacional o internacional de su aprobación condicional de una prueba parcialmente validado, normalmente para un uso durante un periodo limitado en situaciones de emergencia o como base para los acuerdos bilaterales entre países que optan por aceptar los resultados de este ensayo a efectos comerciales.

suficiente para la utilización diagnóstica condicional según establecen las autoridades nacionales. Sin embargo, nunca debe convertirse en un sustituto de una validación de campo completa. Por lo tanto, la validación de referencia de pruebas de diagnóstico sólo puede ofrecer el reconocimiento provisional con previsión de que a continuación se lleve a cabo una validación de campo completa.

Un reconocimiento provisional de una prueba por parte de las autoridades estatales o nacionales reconoce que no se han evaluado las características de rendimiento diagnóstico de la prueba. Como tal, el laboratorio debe desarrollar y seguir un protocolo para la adición y la evaluación de muestras, a medida que estén disponibles, para cumplir con este requisito. Lo ideal es que este proceso se limite a un período de tiempo específico en el cual esta provisión iría dirigida al cumplimiento de las Etapas 2 y 3 de la Validación. Este concepto debe limitarse a las situaciones de emergencia en las que las autoridades consideran esencial la introducción de nuevas pruebas. Puede haber otras situaciones en las que acuerdos comerciales bilaterales, basados en pruebas (por ejemplo, métodos analíticos estándar) que no han sido totalmente validadas en un determinado país, se acepten mutuamente. En casos excepcionales, para las enfermedades raras en las que no existe ninguna otra prueba, se puede permitir el reconocimiento provisional, pero los informes de los resultados deben incluir una declaración del carácter provisional de la validación de la prueba. En todos los casos, debe haber evidencias claras de la estimación preliminar de la DSp y la DSe en base a un pequeño y selecto conjunto de muestras bien caracterizadas que contengan el analito en cuestión.

6.3. Etapa 2 – Rendimiento diagnóstico de la prueba

Sensibilidad y Especificidad Diagnósticas. Los valores de sensibilidad y especificidad diagnóstica constituyen los parámetros más importantes que se establecen durante la validación de una prueba. 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 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 de referencia, 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. El análisis de la curva de eficacia diagnóstica es una herramienta útil para estimar la DSe y la DSp porque evalúa la exactitud global de una prueba diagnóstica cuantitativa teniendo en cuenta los posibles valores de la misma (Greiner y Gardner, 2000; Greiner *et al.*, 2000; Zweig & Campbell, 1993). Este enfoque se describe en detalle en el Anexo 1.1.4.5.

La **sensibilidad diagnóstica** es la proporción de muestras de animales de referencia que se sabe que están infectados que dan un resultado positivo en una prueba.

La **especificidad diagnóstica** es la proporción de muestras de animales de referencia que se sabe que no están infectados que dan un resultado negativo en una prueba.

Debe elegirse un diseño de muestreo que permita estimar la DSe y la DSp. 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, abreviada, 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% muestra una considerable reducción en el número de muestras necesarias. Se requiere un número de muestras bastante grande para lograr una confianza muy alta para la DSe y la DSp cuando se desea un error muy bajo en la estimación. Las limitaciones logísticas y financieras podrían requerir la evaluación de un número de muestras inferior al óptimo. Sin embargo, normalmente no se recomienda reducir los niveles de confianza de la DSe y la DSp por debajo del 90%. 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 (véase el Anexo 1.1.4.5 para más detalle). Por tanto, inicialmente puede ser necesario utilizar una cantidad sub-óptima de muestras. Sin embargo, puede ser muy deseable potenciar la confianza y reducir el error permitido en las estimaciones de la DSe y la DSp añadiendo más muestras a medida que se dispone de ellas.

Tabla 1. Cantidad teórica de muestras procedentes de animales cuyo estado de infección se sabe, necesaria para establecer las estimaciones de sensibilidad (DSe) y especificidad (DSp) diagnósticas con un nivel conocido de confianza

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					
	75%	80%	85%	90%	95%	99%	75%	80%	85%	90%	95%	99%
90%	257	369	475	610	864	1493	41	59	76	98	138	239
92%	210	302	389	466	707	1221	34	48	62	75	113	195
94%	161	232	298	382	542	935	26	37	48	61	87	150
95%	136	196	251	372	456	788	22	31	40	60	73	126
96%	110	158	203	260	369	637	18	25	32	42	59	102
97%	83	119	154	197	279	483	13	19	25	32	45	77
98%	56	80	103	133	188	325	9	13	16	21	30	52
99%	28	41	52	67	95	164	4	7	8	11	15	26

□ Porcentaje de error permitido en la estimación de la DSe o la DSp = 2% en el conjunto de la izquierda y 5% en el conjunto de la derecha. Para calcular el número de muestras necesario para errores del 1%, 3% y 4% en la estimación de la DSe y la DSp, se multiplica el número de muestras del conjunto de la izquierda de la tabla por un factor de 4,0, 0,44 y 0,25, respectivamente.

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

6.3.1. 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 con el agente en cuestión o expuestos al mismo, o que nunca hayan sido infectados ni expuestos. 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.

- i) **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 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 similar a la población de la que procede la muestra.
- ii) **Muestras de referencia positivas:** En general es problemático hallar cantidades suficientes de animales de referencia verdaderamente positivos, comprobados mediante el aislamiento del organismo patógeno. Puede ser necesario recurrir a muestras de animales que se hayan analizado mediante otras pruebas, como los sistemas de detección de ácido nucleico.
- iii) **Muestras de animales de estado desconocido:** en el Anexo 1.1.4.5, Apartado 5.b.i y Apartado 3.a.iii de este capítulo se ofrece una explicación de los modelos de clase latente.

6.3.2 Estado de infección en el animal de referencia

6.3.2.1. El denominado modelo de referencia

El adjetivo “de referencia” se utiliza a menudo para cualificar cualquier método con el que se comparen los demás, pero debe limitarse a los métodos o combinaciones de métodos que invariablemente clasifiquen de forma correcta a los animales en infectados/expuestos o no infectados. Algunos métodos de aislamiento presentan en sí mismos problemas de repetibilidad y sensibilidad analítica, de modo que no son verdaderos métodos de referencia, en concreto por el análisis de muestras supuestamente negativas. Cuando el denominado estándar de referencia es imperfecto, que es el caso habitual de la mayoría de pruebas ante-mortem, las estimaciones de la DSe y de la DSp de la prueba candidata podrían resultar comprometidas porque el error de las

estimaciones obtenidas por comparación con el estándar de referencia se arrastra a las estimaciones correspondientes al prueba candidata. De hecho, al utilizar pruebas de referencia imperfectas, las estimaciones de la DSe y la DSp del rendimiento de la prueba candidata serán imperfectas y a menudo se sobreestimarán.

Las pruebas de DAN podrían ser más sensibles y específicas que los existentes métodos de referencia, lo cual convierte al método de referencia establecido en no adecuado para su uso con fines de comparación. Si la DAN es más sensible que el método de referencia, la aparente menor especificidad en el método de referencia resultará desorientadora. Este problema se puede resolver en parte evaluando el origen de la muestra, su historial clínico y la secuenciación de todos los productos de la PCR para confirmar la identidad del analito.

6.3.2.2. Modelos de clase latente

Los modelos de clase latente (Branscum *et al.*, 2005; Enoe *et al.*, 2000; Georgiadis *et al.*, 2003; Hui & Walter, 1980) no se basan en la suposición de una prueba de referencia perfecta sino que estiman la exactitud de la prueba candidata y del método de referencia con los resultados de ambas pruebas. Dado que estos modelos estadísticos son complejos y requieren suposiciones críticas, se precisaría ayuda estadística para contribuir a orientar el análisis y describir el muestreo en la población o poblaciones estudiadas, las características de otras pruebas incluidas en el análisis, la correcta elección de modelo y los métodos de estimación en base a bibliografía revisada por pares (véase el Anexo 1.1.4.5 sobre consideraciones estadísticas para más detalle).

6.3.3. Animales de referencia infectados experimentalmente o vacunados

Los sueros obtenidos 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. 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. Solo es aceptable el muestreo de animales experimentales realizado en un solo momento. Además, 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 experimentales son ejemplos de variables que pueden inducir a error cuando se extrapolan las estimaciones de sensibilidad y especificidad diagnósticas a la población estudiada. Por tales razones, la validación de una prueba no debe basarse sólo en muestras de animales experimentales.

6.3.4. Determinación de los umbrales (puntos de corte)

Para obtener estimaciones de la DSe y la DSp de la prueba candidata, en primer lugar los resultados de la prueba deben convertirse en categóricos (positivo, negativo o intermedio). Esto se consigue insertando uno o dos puntos de corte (umbrales o límites de decisión) en la escala continua de resultados de la prueba. Los puntos de corte escogidos deben reflejar el objetivo de la prueba y su aplicación, y deben respaldar la DSe y la DSp de la prueba requeridas. 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.*, 1994; 2000; Jacobson, 1998; Zweig & Campbell, 1993; y el Anexo 1.1.4.5 sobre consideraciones estadísticas). Si se produce una superposición considerable en las distribuciones de los valores de las pruebas con animales infectados y no infectados, resulta difícil 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). Los valores que caigan entre estos percentiles pueden clasificarse como intermedios (véase el recuadro), y requerirían ser analizados mediante una prueba confirmativa, un segundo análisis para detectar la seroconversión, o una secuenciación con fines de identificación.

Umbral y punto de corte se consideran sinónimos. Un punto de corte es el valor de la prueba escogido para diferenciar entre resultados negativos y positivos en una escala continua de valores de dicha prueba.

Intermedio, inconcluyente, sospechoso o ambiguo son adjetivos utilizados de forma indiferente para cualificar un intervalo de valores situado entre los puntos de corte positivo y negativo.

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. Se explican alternativas en el Apartado 6.2.7. sobre la aceptación provisional de una prueba durante la recogida de datos para mejorar las estimaciones de la DSe y de la DSp.

6.3.5. Cálculo de la DSe y de la DSp en base a los resultados del análisis de sueros de referencia

Un método habitual de determinación de las estimaciones de la DSe y la DSp es analizar las muestras de referencia con la nueva prueba, 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 decidió que la DSe y la DSp estimadas para la nueva prueba debían ser del 97% y del 99%, respectivamente, con una confianza deseada del 95% para 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. A continuación, se analizaron las muestras mediante la nueva prueba. La Tabla 2 es un conjunto hipotético de resultados y los cálculos de las estimaciones de la DSe y de la DSp se basan en las muestras analizadas.

Tabla 2. 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	7
	Negativo	9	88

VP	FP
FN	VN

Sensibilidad diagnóstica*	Especificidad diagnóstica*
$VP/(VP + FN)$	$VN/(VN + FP)$
96,7% (94,0 – 98,5%)**	92,0% (84,3 – 96,7%)**

*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 Anexo 1.1.4.5 para información sobre límites de confianza)

En este ejemplo, las estimaciones de la DSe son las previstas, pero la DSp es muy inferior al previsto 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.

6.4. Etapa 3 – Estimaciones de la reproducibilidad y del aumento de la repetibilidad

La reproducibilidad es una importante medida de la precisión de una prueba cuando se utiliza en varios laboratorios situados en regiones o países distintos utilizando exactamente la misma prueba (protocolo, reactivos y controles). Cada uno de al menos tres laboratorios analizan el mismo conjunto de un mínimo de 20 muestras (sin que los técnicos sepan cuál es cada una), con idénticas alícuotas para cada laboratorio (véase el Anexo 1.1.4.7 sobre conjuntos de muestras). Este ejercicio también genera datos

La **solidez** es una medida de la capacidad de la prueba de no resultar afectado por cambios sustanciales o sustituciones en las condiciones de la prueba previstas para cuando se lleve a cabo en varios laboratorios.

La **reproducibilidad** es la capacidad de un método analítico de proporcionar resultados constantes, determinados mediante estimaciones de la precisión, cuando se aplique a alícuotas de las mismas muestras analizadas en distintos laboratorios.

preliminares sobre efectos no aleatorios atribuibles a la utilización de la prueba en otros laboratorios - también se denomina solidez de la prueba. 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 Anexo 1.1.4.4, sobre Medición de la Incertidumbre, se explica en mayor detalle este tema y su aplicación).

6.5. Etapa 4 – Implementación del programa

6.5.1. Interpretación de los resultados de la prueba

Valores predictivos de los resultados de la prueba. Los resultados de una prueba son principalmente útiles cuando las inferencias que se realizan a partir de ellos son exactas. Los valores predictivos de los resultados de la prueba deben basarse en la verdadera prevalencia de exposición/infección en la población estudiada. En las pruebas de cribado que se utilizan en la vigilancia de una población “libre de la enfermedad”, los resultados falsos positivos constituyen un importante problema. Por ejemplo, una prueba puede tener una fiabilidad impecable (por ejemplo, una alta precisión y una alta exactitud, con un 99% de DSe y un 99,9% de DSp), pero si la prevalencia de la enfermedad es próxima a cero, y la prueba tiene un falso positivo por cada 1.000 animales analizados, las inferencias realizadas a partir de los falsos positivos suponen un problema (en Jacobson, 1998 se exponen tablas del VP de distintas estimaciones de la DSe y la DSp). De forma similar, si al utilizar la prueba para detectar una enfermedad muy virulenta se obtiene un falso negativo por cada 100 animales analizados, las inferencias realizadas a partir de los falsos negativos podrían tener consecuencias devastadoras. Esto ilustra la importancia crucial de elegir umbrales de diagnóstico que sean adecuados para la aplicación en cuestión. Los umbrales deben escogerse para minimizar el efecto de los falsos positivos y/ falsos negativos en los valores predictivos de la prueba dada su aplicación y la prevalencia de la exposición/infección en la población estudiada. Tal vez también sea prudente disponer de pruebas confirmativas muy específicas para determinar si los animales que dan positivo en la prueba de cribado son verdaderos o falsos positivos.

Valor predictivo (VP) de los resultados positivos o negativos de la prueba. El VP+ es la probabilidad de que un animal que da un resultado positivo haya estado expuesto o infectado. El VP es la probabilidad de que un animal que da un resultado negativo no haya estado expuesto ni esté infectado.

En el caso de las pruebas de detección de ácido nucleico, podría ser necesario confirmar los resultados positivos de la DAN mediante análisis de la secuencia del producto amplificado (un ejemplo de pruebas que contribuyen a resolver errores debidos a una secuencia de acceso o punto de unión del cebador inespecíficos).

6.5.2 Reconocimiento internacional

Tradicionalmente, la OIE ha reconocido internacionalmente pruebas cuando están diseñadas como 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 que han superado el proceso de certificación, el paso final es incluir la prueba en el Registro de la OIE. Las pruebas incluidas en el Registro están certificadas como adecuadas para un fin concreto si han superado las etapas de Validación 1, 2 y 3. El Registro tiene por objetivo proporcionar a los posibles usuarios de la prueba una fuente informada e imparcial de información sobre la prueba y sus características de rendimiento para un fin determinado. Este Registro puede consultarse en la página web de la OIE: <http://www.oie.int/en/our-scientific-expertise/certification-of-diagnostic-tests/background-information/>

6.5.3. Utilización de la prueba

La demostración definitiva de la utilidad de una prueba analítica es su aplicación con éxito en otros laboratorios y su inclusión en programas nacionales, regionales y/o internacionales. Los laboratorios de referencia desempeñan un papel crucial en este proceso. En el avance natural de las mejoras diagnósticas y/o tecnológicas, nuevas pruebas se convertirán en el nuevo método de referencia frente al cual se compararán los 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 solidez cuando la prueba se transfiere del laboratorio al campo, tanto para su uso en laboratorios locales como en aplicaciones a pie de granja. 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 solidez (que principalmente deriva de las estimaciones de la reproducibilidad).

6.6. Seguimiento del rendimiento de la prueba tras la validación inicial

6.6.1. Seguimiento de la prueba

Una prueba validada y que se utilice de forma sistemática precisa un seguimiento constante de su repetibilidad mediante controles durante el proceso, con el fin de evaluar posibles cambios de la precisión y la exactitud de la prueba a lo largo del tiempo. Puede realizarse un seguimiento de estos posibles cambios gráficamente trazando valores control en diagramas de control. 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. También es fundamental llevar a cabo una posterior evaluación continua del rendimiento de la prueba y suele realizarse mediante evaluaciones de la precisión, la exactitud y de tendencias atípicas utilizando diagramas de control. La reproducibilidad se evalúa mediante programas de control de calidad externos, como el análisis de la eficiencia.

Diagrama de control - Es una representación gráfica de los datos de varias mediciones de una muestra control en distintas ejecuciones de la prueba a lo largo

6.6.2. Modificaciones en el diagnóstico - 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 afrontar cambios en los analitos estudiados (es decir, modificaciones en la prueba para ajustar el rendimiento diagnóstico) o bien pueden ser necesarias modificaciones técnicas para mejorar la eficiencia o la relación coste-eficacia de la prueba.

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 un virus presentes en animales de distintas regiones geográficas, tienen distintas secuencias de acceso o puntos de unión de los cebadores, lo cual implica tener que revalidar la prueba. Esto es especialmente cierto en el caso de los sistemas DAN y es muy frecuente que se produzcan mutaciones puntuales en muchos agentes infecciosos (es decir, 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 en 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 las estimaciones de la DSe de la prueba no resultan comprometidas.

Puede producirse una situación similar con la entrada de nuevos linajes víricos en países o regiones en que ese linaje vírico no existía previamente. En estas circunstancias, las pruebas de DAN existentes que no iban dirigidas a estos nuevos linajes pueden precisar una modificación para incluir cebadores o sondas dirigidos a estos nuevos analitos. Lo mismo sería aplicable a la tipificación de sueros utilizados en pruebas de neutralización de virus.

6.6.2.1. 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 los resultados de la prueba. 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 los resultados del procedimiento modificado y del método original validado se determina que son comparables en un experimento basado en un criterio pre-especificado, la prueba modificada sigue considerándose validada para el fin deseado. En el Anexo 1.1.4.6 se describen experimentos que son adecuados para comprobar la comparabilidad y en el Anexo 1.1.4.7 se indican conjuntos de muestras de referencia.

Una evaluación de la comparabilidad puede o no ser adecuada si la prueba se aplica a una matriz de muestra diferente, por ejemplo, en el caso de que la prueba esté validada en sangre y se utilice en otro tejido de la especie de la cual se ha extraído la muestra. Puede ser necesaria una revalidación si la prueba está diseñada para su uso en una nueva especie.

6.6.2.2. Sustitución de reactivos agotados

Cuando un reactivo, como una muestra control, está a punto de agotarse, resulta necesario preparar y analizar repetidamente un reactivo de repuesto antes de que el primero se agote. 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.

6.6.3. 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 de animales cuyo estado de infección se conozca. Esto 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. Deben incluirse nuevos datos cada año en los dossier de pruebas correspondientes.

BIBLIOGRAFÍA

- BALLAGI-PORDÁNY A. & BELÁK S. (1996). The use of mimics as internal standards to avoid false negatives in diagnostic PCR. *Mol. Cell. Probes*, **10**, 159–164.
- BELÁK S. (2005). The molecular diagnosis of porcine viral diseases: a review. *Acta Vet. Hung.*, **53**, 113–124. (Review).
- BELÁK S. (2007). Molecular diagnosis of viral diseases, present trends and future aspects. A view from the OIE Collaborating Centre for the Application of Polymerase Chain Reaction Methods for Diagnosis of Viral Diseases in Veterinary Medicine. *Vaccine*, **25**, 5444–5452.
- BELÁK S. & THORÉN P. (2001). Molecular diagnosis of animal diseases. *Expert Rev. Mol. Diagn.*, **1**, 434–444.
- BRANSCUM A.J., GARDNER I.A. & JOHNSON W.O. (2005). Estimation of diagnostic-test sensitivity and specificity through Bayesian modeling. *Prev. Vet. Med.*, **68**, 145–163.
- BURNS M.J., NIXON G.J., FOY C.A. & HARRIS N. (2005). Standardisation of data from real-time quantitative PCR methods - evaluation of outliers and comparison of calibration curves. *BMC Biotechnol.*, **5**, 31–44.
- BUSTIN S.A. (2005). Real-time, fluorescence-based quantitative PCR: a snapshot of current procedures and preferences. *Expert Rev. Mol. Diagn.*, **5**, 493–498.
- 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** (3), 913–935.
- DEJAEGER B. & VANDER HEYDEN Y. (2006). Robustness tests. *LCGC Europe*, **19** (7) online at <http://www.lcgeurope.com/lcgeurope/content/printContentPopup.jsp?id=357956>
- 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 (<http://www.aapsj.org>))
- 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.
- GREINER M., FRANKE C.R., BOHNING D. & SCHLATTMANN P. (1994). Construction of an intrinsic cut-off value for the sero-epidemiological study of *Trypanosoma evansi* infections in a canine population in Brazil: a new approach towards unbiased estimation of prevalence. *Acta Trop.*, **56**, 97–109.
- GREINER M. & GARDNER I. (2000). Epidemiologic issues in the validation of veterinary diagnostic tests. *Vet. Prev. 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.

HUGGETT J., DHEDA K., BUSTIN S. & ZUMLA A. (2005). Real-time RT-PCR normalisation; strategies and considerations. *Genes Immun.*, **6**, 279–284. (Review).

HUI S.L. & WALTER S.D. (1980). Estimating the error rates of diagnostic tests. *Biometrics*, **36**, 167–171.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION/INTERNATIONAL ELECTROTECHNICAL COMMISSION (ISO/IEC) (2005). ISO/IEC 17025:2005, General requirements for the competence of testing and calibration laboratories.

JACOBSON R.H. (1998). Validation of serological assays for diagnosis of infectious diseases. *Rev. sci. tech. Off. int. Epiz.*, **17**, 469–486.

LAUERMAN L.H. (2004). Advances in PCR technology. *Anim. Health Res. Rev.*, **5**, 247–248.

LOUIE M., LOUIE L. & SIMOR A.E. (2000). The role of DNA amplification technology in the diagnosis of infectious diseases. *CMAJ.*, **163**, 301–309.

THOMPSON, M., ELLISON, S. & WOOD, R. (2002). Harmonized guidelines for single-laboratory validation of methods of analysis. *Pure Appl. Chem.*, **75** (5), 835–855.

INTERNATIONAL CONFERENCE ON HARMONISATION (ICH) (2005). Validation of Analytical Procedures: Text and Methodology (Q2[R1]). Tripartite Guideline of the International Conference on Harmonisation for Technical Requirements for Registration of Pharmaceuticals for Human Use, pp 1–13
(Available on-line at: http://bioforum.org.il/uploads/Editor/karen/q2_r1_step4.pdf)

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** (8), 1381–1386.

WORLD ORGANISATION FOR ANIMAL HEALTH (OIE) (2008). OIE Standard for Management and Technical Requirements for Laboratories Conducting Tests for Infectious Diseases. *In: OIE Quality Standard and Guidelines for Veterinary Laboratories: Infectious Diseases*. OIE, Paris, France, 1–31.

YOU DEN W. & STEINER E. (1987). Statistical Manual of the AOAC. AOAC International. 96 Pages. 5th Printing (ISBN 0-935584-15-3).

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

*
* *

ANEXOS AL CAPÍTULO SOBRE VALIDACIÓN - EN ESTUDIO

Anexo 1.1.4.1. Desarrollo y optimización de pruebas de detección de anticuerpos

Anexo 1.1.4.2. Desarrollo y optimización de pruebas de detección de anticuerpos

Anexo 1.1.4.3. Desarrollo y optimización de pruebas de Detección de Ácido Nucleico (DAN)

Anexo 1.1.4.4. Medición de la Incertidumbre

Anexo 1.1.4.5. Enfoques Estadísticos de la Validación

Anexo 1.1.4.6. Comparabilidad de las pruebas tras cambios menores en un método analítico validado

Anexo 1.1.4.7. Conjuntos de Muestras y Valores de Referencia - Selección y Utilización