

comisión del codex alimentarius



ORGANIZACIÓN DE LAS NACIONES
UNIDAS PARA LA AGRICULTURA
Y LA ALIMENTACIÓN

ORGANIZACIÓN
MUNDIAL
DE LA SALUD



OFICINA CONJUNTA: Viale delle Terme di Caracalla 00100 ROMA Tel: 39 06 570151 www.codexalimentarius.net Email: codex@fao.org Facsimile: 39 06 5705 493

Tema 8a) del programa

CX/MAS 02/10

PROGRAMA CONJUNTO FAO/OMS SOBRE NORMAS ALIMENTARIAS COMITÉ DEL CODEX SOBRE MÉTODOS DE ANÁLISIS Y TOMA DE MUESTRAS

24ª reunión

Budapest, Hungría, 18 – 22 de noviembre de 2002

VALIDACIÓN POR UN SOLO LABORATORIO

EXAMEN DE LAS DIRECTRICES ARMONIZADAS DE LA UIQPA PARA
LA VALIDACIÓN INTERNA DE MÉTODOS DE ANÁLISIS

Antecedentes

El Comité sobre Métodos de Análisis y Toma de Muestras acordó en su 21ª reunión comenzar a trabajar en la validación interna de métodos de análisis. En su 22ª reunión observó que la Unión Internacional de Química Pura y Aplicada (UIQPA) estaba elaborando unas *Directrices armonizadas para la validación interna de métodos de análisis* y acordó examinarlas con más detenimiento cuando estuviera disponible el siguiente borrador (ALINORM 99/23, párrafos 47-51). En su 23ª reunión observó que las Directrices no estaban aún publicadas en su formato final y debatió el enfoque general de la validación por un solo laboratorio para los fines del Codex, teniendo en cuenta los trabajos en curso en este sector (ALINORM 01/23, párrafos 65-84).

El Comité reconoció de forma general que la validación por un solo laboratorio podía utilizarse para los fines del Codex y acordó que en la siguiente reunión se estudiaría la inclusión en el Manual de Procedimiento de un texto que tratara esta cuestión de forma específica. El Comité acordó asimismo que la delegación de los Países Bajos revisara el documento sobre requisitos para la validación por un solo laboratorio y que la delegación del Reino Unido elaborara otro documento sobre la validación de métodos basándose en los resultados de planes de ensayos de aptitud. Estos documentos se identificarán, respectivamente, como CX/MAS 02/11 y CX/MAS 02/12 en relación con los **temas 8b) y 8c) del programa**. Se informará también al Comité acerca de las novedades relativas a la aplicación de la validación por un solo laboratorio en el Comité sobre Residuos de Plaguicidas y en el Comité sobre Residuos de Medicamentos Veterinarios en los Alimentos, en relación con el **tema 2 del programa**.

El Comité acordó que en su siguiente reunión examinaría la versión publicada de las *Directrices armonizadas para la validación interna de métodos de análisis* de la UIQPA, con el propósito de adoptarlas por referencia. A la fecha, el título de las Directrices utilizaba la expresión "validación interna" y así se hizo constar en el programa. Posteriormente, se modificó el título de la Directrices finales y la anterior expresión se sustituyó por "validación por un solo laboratorio"; esta fue también la expresión recomendada por el Comité en su última reunión.

Se adjunta, para su examen y posibles observaciones, la versión final de las *Directrices armonizadas para la validación de métodos de análisis por un solo laboratorio* de la UIQPA publicadas recientemente. De

conformidad con su decisión anterior, el Comité examinará si debe recomendar o no que la Comisión adopte las *Directrices* por referencia para los fines del Codex.

Los Gobiernos y organizaciones internacionales que deseen presentar sus observaciones deberán hacerlo por escrito (preferentemente mediante correo electrónico) a: Secretario de la Comisión del Codex Alimentarius Commission, Programa Conjunto FAO/OMS sobre Normas Alimentarias, FAO, Viale delle Terme di Caracalla, 00100 Roma, Italia, y remitir una copia a: Dr. Mária Váradi, Central Food Research Institute (KÉKI), Herman Ottó út 15, H-1022 Budapest, Fax: +361 212 9853 o 361 355 8928, correo electrónico: m.varadi@mail.cfri.hu, **antes del 15 de octubre de 2002.**

Pure Appl. Chem., Vol. 74, No. 5, pp. 835–855, 2002.

© 2002 UIQPA

UNIÓN INTERNACIONAL DE QUÍMICA PURA Y APLICADA

DIVISIONES DE QUÍMICA ANALÍTICA, APLICADA, CLÍNICA, INORGÁNICA Y FÍSICA
GRUPO DE TRABAJO INTERDIVISIONAL PARA LA ARMONIZACIÓN DE SISTEMAS DE GARANTÍA
DE LA CALIDAD PARA LABORATORIOS DE ANÁLISIS*

DIRECTRICES ARMONIZADAS PARA LA VALIDACIÓN DE MÉTODOS DE ANÁLISIS POR UN SOLO LABORATORIO (Informe técnico de la UIQPA)

Resultado del Simposio sobre armonización de sistemas de garantía de la calidad para laboratorios de análisis,
Budapest, Hungría, 4 y 5 de noviembre de 1999, auspiciado por la UIQPA, la ISO y por AOAC International

Preparado para publicación por

MICHAEL THOMPSON¹, STEPHEN L. R. ELLISON² Y ROGER WOOD^{3‡}

¹ *Department of Chemistry, Birkbeck College (University of London), London WC1H 0PP, Reino Unido*

² *Laboratory of the Government Chemist, Queens Road, Teddington, Middlesex TW11 0LY, Reino Unido*

³ *Food Standards Agency, c/o Institute of Food Research, Norwich Research Park, Colney, Norwich, NR4 7UA,
Reino Unido*

* En el período 1997–2000 el Grupo de Trabajo estuvo constituido por los siguientes miembros:

Presidente: A. Fajgelj, 1997–, (OIEA, Austria); **Miembros:** K. Bergknut (Noruega); Carmen Camara (España); K. Camman (Alemania); Jytte Molin Christensen (Dinamarca); S. Coates (AOAC Int., Estados Unidos de América); P. De Bièvre (Bélgica); S. L. R. Ellison (Reino Unido); T. Gills (Estados Unidos de América); J. Hlavay (Hungría); D. G. Holcombe (Reino Unido); P. T. Holland (Nueva Zelanda); W. Horwitz (Estados Unidos de América); A. Kallner (Suecia); H. Klich (Alemania); E. A. Maier (Bélgica); C. Nieto De Castro (Portugal); M. Parkany (Suiza); J. Pauwels (Bélgica); M. Thompson (Reino Unido); M. J. Vernengo (Argentina); R. Wood (Reino Unido).

‡ Autor corresponsal

Se permite la republicación o la reproducción de este informe o su archivamiento y/o difusión por medios electrónicos sin necesidad de solicitar permiso formal a la UIQPA, a condición de que se declare, en un lugar claramente visible, con referencia completa de la fuente, la autoría de la UIQPA, indicando su nombre junto con la inclusión del símbolo de derecho de autor © y el año de publicación. La publicación de una traducción a otro idioma está sujeta a la condición adicional de la aprobación previa del correspondiente organismo nacional adherido a la UIQPA.

M. THOMPSON *et al.***Directrices armonizadas para la validación de métodos de análisis por un solo laboratorio****(Informe técnico de la UIQPA)**

Resumen: La validación de métodos es una de las medidas reconocidas universalmente como parte necesaria de un sistema integral de garantía de la calidad en el ámbito de la química analítica. La ISO, la UIQPA y AOAC International han colaborado anteriormente para elaborar protocolos o directrices consensuados sobre diseño, organización e interpretación de estudios de funcionamiento de métodos [1], sobre ensayos de aptitud de laboratorios de análisis (químico) [2], sobre control de calidad interno en laboratorios de química analítica [3] y sobre el uso de la información de recuperación en las mediciones analíticas [4]. El grupo de trabajo que elaboró estos protocolos o directrices ha recibido ahora el encargo de la UIQPA de elaborar directrices sobre la validación de métodos de análisis por un solo laboratorio. Estas directrices constituyen recomendaciones de mínimos sobre los procedimientos que deben aplicarse para asegurar la validación apropiada de métodos de análisis.

Se ha examinado un borrador de las directrices en un Simposio Internacional sobre la Armonización de Sistemas de Garantía de la Calidad en Laboratorios de Química, cuyas deliberaciones han sido publicadas por la Royal Society of Chemistry del Reino Unido.

ÍNDICE

1.	INTRODUCCIÓN	837
1.1	Antecedentes	
1.2	Protocolos, normas y directrices existentes	
2.	DEFINICIONES Y TERMINOLOGÍA	838
2.1	Consideraciones generales	
2.2	Definiciones utilizadas únicamente en la presente guía	
3.	VALIDACIÓN DE MÉTODOS, INCERTIDUMBRE Y GARANTÍA DE LA CALIDAD	839
4.	PRINCIPIOS FUNDAMENTALES DE LA VALIDACIÓN DE MÉTODOS	839
4.1	Especificación y ámbito de aplicación de la validación	
4.2	Comprobación de hipótesis	
4.3	Fuentes de error en el análisis	
4.4	Efectos del método y del laboratorio	
5.	REALIZACIÓN DE ESTUDIOS DE VALIDACIÓN	843
6.	ALCANCE DE LOS ESTUDIOS DE VALIDACIÓN	844
7.	RECOMENDACIONES	845
	APÉNDICE A: NOTAS SOBRE LOS REQUISITOS PARA EL ESTUDIO DE LAS CARACTERÍSTICAS DE FUNCIONAMIENTO DE LOS MÉTODOS	845
A1.	Aplicabilidad	
A2.	Selectividad	
A3.	Calibración y linealidad	
	<i>A3.1 Linealidad e intercepción</i>	
	<i>A3.2 Ensayo del efecto general de la matriz</i>	
	<i>A3.3 Procedimiento de calibración final</i>	

- A4. Veracidad
 - A4.1 Estimación de la veracidad
 - A4.2 Condiciones para la realización de experimentos de estimación de la veracidad
 - A4.3 Valores de referencia para experimentos de estimación de la veracidad
 - A4.3.1 Materiales de referencia certificados (MRC)
 - A4.3.2 Materiales de referencia
 - A4.3.3 Utilización de un método de referencia
 - A4.3.4 Adición (*spiking*) y recuperación
- A5. Precisión
- A6. Recuperación
- A7. Intervalo
- A8. Límite de detección
- A9. Límite de determinación o límite de cuantificación
- A10. Sensibilidad
- A11. Robustez
- A12. Aptitud para los fines
- A13. Variación de la matriz
- A14. Incertidumbre de la medición

APÉNDICE B. CONSIDERACIONES ADICIONALES PARA LA ESTIMACIÓN DE LA INCERTIDUMBRE EN ESTUDIOS DE VALIDACIÓN	854
B1. Análisis de la sensibilidad	
B2. Juicio profesional	
REFERENCIAS	855

1. INTRODUCCIÓN

1.1 Antecedentes

Para el cumplimiento de normas nacionales e internacionales en todos los campos de análisis se necesitan métodos de análisis fiables. La comunidad internacional reconoce, en consecuencia, que los laboratorios deben adoptar medidas adecuadas para asegurar que son capaces de proporcionar y que proporcionan datos de la calidad exigida. Dichas medidas comprenden:

- la utilización de métodos de análisis validados;
- la utilización de procedimientos de control de calidad interno;
- la participación en programas de ensayos de aptitud y
- la acreditación del cumplimiento de una norma internacional, habitualmente la Norma ISO/IEC 17025.

Se debe señalar que la acreditación ISO/IEC 17025 aborda de forma específica la aplicación de un sistema de rastreabilidad de las mediciones, además de requerir otras condiciones técnicas y de gestión, entre las que se incluyen todas las mencionadas en la relación anterior.

La validación de métodos es por consiguiente un componente esencial de las medidas que debe aplicar un laboratorio para obtener datos analíticos fiables. El Grupo de Trabajo Interdivisional sobre la Armonización de Sistemas de Garantía de la Calidad para Laboratorios de Análisis de la UIQPA ha abordado anteriormente otros aspectos de este contexto; en concreto, ha elaborado protocolos o directrices sobre estudios (en colaboración) de funcionamiento de los métodos [1], ensayos de aptitud [2] y control de calidad interno [3].

En algunos sectores, principalmente en el análisis de alimentos, la exigencia de disponer de métodos que han sido “validados plenamente” es una prescripción de ley [5,6]. Se considera habitualmente que la validación “plena” de un método de análisis comprende un examen de las características del método en un estudio del funcionamiento del método entre laboratorios (conocido también como estudio en colaboración o ensayo en colaboración). Se han establecido protocolos aceptados internacionalmente para la validación “plena” de un método de análisis mediante un ensayo en colaboración, entre los que destacan el Protocolo armonizado internacional [1] y el procedimiento de la ISO [7]. En estos protocolos o normas se exigen, para la validación plena del método de análisis, la participación en el ensayo en colaboración de un número mínimo de laboratorios

y materiales de ensayo. Sin embargo, no siempre es práctico o necesario realizar una validación plena de los métodos de análisis. En estos casos, puede ser apropiado proceder a una “validación de métodos por un solo laboratorio”.

La validación de métodos por un solo laboratorio es apropiada en diversas circunstancias, como las siguientes:

- para asegurar la viabilidad del método antes de emprender un costoso ensayo en colaboración formal;
- para demostrar la fiabilidad de métodos de análisis si no se dispone de datos de ensayos en colaboración o cuando no sea factible realizar un ensayo en colaboración formal y
- para asegurarse de que los métodos validados corrientes se aplican correctamente.

Cuando se prevé la caracterización interna de un método, es importante que el laboratorio determine y acuerde con su cliente qué características concretas se han de evaluar. No obstante, en ciertas situaciones estas características pueden estar fijadas por la ley (por ejemplo, residuos de medicamentos veterinarios en alimentos y plaguicidas en los sectores alimentarios). El ámbito de la evaluación realizada por un laboratorio debe ajustarse a los requisitos normativos.

No obstante, en algunos sectores de análisis, numerosos laboratorios utilizan el mismo método de análisis para determinar compuestos químicos estables en matrices definidas. Debe tenerse en cuenta que si se puede proporcionar a estos laboratorios un método estudiado en colaboración adecuado, los costos del ensayo en colaboración para validar dicho método pueden estar plenamente justificados. La utilización de un método estudiado en colaboración reduce considerablemente la inversión que debe realizar un laboratorio en una labor exhaustiva de validación antes de aplicar el método de forma sistemática. Si un laboratorio utiliza un método estudiado en colaboración, que se ha comprobado que es adecuado para la finalidad prevista, únicamente debe demostrar que puede alcanzar las características de funcionamiento especificadas en el método. Esta verificación del uso correcto de un método es mucho menos costosa que una validación plena por un solo laboratorio. El costo total para la comunidad de analistas de la validación de un determinado método mediante un ensayo en colaboración y la posterior verificación de sus características de funcionamiento en los laboratorios interesados en utilizar el método es a menudo menor que si numerosos laboratorios realizaran una validación del mismo método de forma independiente.

1.2 Protocolos, normas y directrices existentes

Se han elaborado diversos protocolos y directrices [8–19] sobre la validación de métodos y la incertidumbre, entre los que destacan los descritos en los siguientes documentos de AOAC Internacional, la Conferencia Internacional sobre Armonización (ICH) y Eurachem:

- el manual de estadística de la AOAC, que incluye una orientación sobre los estudios en un solo laboratorio anteriores a la realización de ensayos en colaboración [13];
- el texto [15] y la metodología [16] de validación de procedimientos de análisis de la ICH que establecen los requisitos mínimos que deben cumplir los estudios de validación de los ensayos que se utilizan para respaldar las propuestas de aprobación de medicamentos;
- *The Fitness for Purpose of Analytical Methods: A Laboratory Guide to Method Validation and Related Topics* (1998) [12] y
- *Quantifying Uncertainty in Analytical Measurement* (2000) [9].

La validación de métodos también examinó a fondo, en diciembre de 1997, en una consulta de expertos FAO/OIEA sobre validación de métodos de análisis para controles alimentarios, de la que se ha publicado un informe [19].

Las presentes directrices reúnen los principios científicos esenciales de los anteriores documentos, con objeto de proporcionar información que ha sido aceptada por la comunidad internacional y, lo que es más importante, de señalar el camino que debe seguirse para establecer prácticas óptimas de validación de métodos en un solo laboratorio.

2 DEFINICIONES Y TERMINOLOGÍA

2.1 Consideraciones generales

Los términos empleados en el presente documento se ajustan a las definiciones de la ISO y la UIQPA vigentes. Los siguientes documentos contienen definiciones pertinentes:

- IUPAC: *Compendium of Chemical Terminology*, segunda edición, A. D. McNaught y A. Wilkinson, 1997 (conocido en inglés como “*the Gold Book*” en atención a Victor Gold, promotor de la primera edición).
- ISO: *Vocabulario internacional de términos básicos y generales de metrología*, 1993

2.2 Definiciones utilizadas únicamente en la presente guía

Incertidumbre relativa: incertidumbre expresada en términos de la desviación típica relativa.

Intervalo validado: la parte del intervalo de concentración de un método de análisis que ha sido validado.

3. VALIDACIÓN DE MÉTODOS, INCERTIDUMBRE Y GARANTÍA DE LA CALIDAD

Para la validación de métodos se emplea un conjunto de ensayos que comprueban las hipótesis en las que se basa el método de análisis y determinan y documentan las características de funcionamiento de un método, demostrando así su aptitud para un determinado fin analítico. Las características de funcionamiento típicas de los métodos de análisis son: aplicabilidad, selectividad, calibración, veracidad, precisión, recuperación, intervalo operativo, límite de cuantificación, límite de detección, sensibilidad y robustez. A estas características se pueden añadir la incertidumbre de la medición y la aptitud para los fines.

En rigor, más que a un “método de análisis” la validación debería referirse a un “sistema de análisis” que comprende un protocolo definido del método, un intervalo definido de concentración del analito y un tipo especificado de material de ensayo. Para los fines del presente documento, se entenderá que la expresión “validación de métodos” se refiere a un sistema de análisis completo. Para referirse al procedimiento de análisis propiamente dicho, se utilizará la expresión “el protocolo”.

En el presente documento, la validación de métodos se considera una actividad diferente de actividades corrientes como el control de calidad interno (CCI) o los ensayos de aptitud. La validación de un método, que se realiza una sola vez o a intervalos relativamente poco frecuentes durante la duración útil del mismo, nos indica su funcionamiento futuro esperado. El CCI nos indica la evolución temporal del funcionamiento del método. En consecuencia, el CCI se aborda como una actividad independiente en el programa de armonización de la UIQPA [3].

Las características cuantitativas de interés en la validación de métodos son las relativas a la exactitud probable del resultado obtenido. Por lo tanto, se puede afirmar de forma general que la validación de métodos consiste esencialmente en la estimación de la incertidumbre de la medición. A lo largo de los años, se ha establecido como práctica habitual la representación, para fines de validación, de las diversas características del funcionamiento de los métodos por referencia a los elementos independientes indicados anteriormente y las presentes directrices se ajustan en gran medida a este patrón. No obstante, dado que la incertidumbre de la medición se utiliza cada vez más como indicador clave tanto de la aptitud para los fines como de la fiabilidad de los resultados, cada vez más profesionales de la química analítica acometerán la validación de las mediciones para respaldar la estimación de la incertidumbre y algunos profesionales querrán hacerlo de forma inmediata. En consecuencia, en el Apéndice A se aborda concisamente la cuestión de la incertidumbre de la medición como característica de funcionamiento de los métodos de análisis y en el Apéndice B se proporciona una orientación adicional sobre algunos procedimientos no abordados en el resto del documento.

4. PRINCIPIOS FUNDAMENTALES DE LA VALIDACIÓN DE MÉTODOS

4.1 Especificación y ámbito de aplicación de la validación

La validación se aplica a un protocolo definido, para la determinación de un analito e intervalo de concentraciones especificados, en un tipo determinado de material de ensayo, empleado para un fin determinado. En general, mediante la validación se debe comprobar que el funcionamiento del método es adecuado para su fin, para toda la gama de materiales de ensayo y en los intervalos de concentración de analito a los que se aplica.

En consecuencia, antes de acometer la validación, se deben especificar completamente estas características y se deben declarar los criterios de aptitud para los fines.

4.2 Comprobación de hipótesis

Además de proporcionar cifras de funcionamiento indicativas de la aptitud para los fines y que han llegado a dominar el uso de los datos de validación en la práctica, los estudios de validación sirven para comprobar de forma objetiva las hipótesis en las que se basa un método de análisis. Por ejemplo, si el cálculo de un resultado se basa en una función de calibración que es una simple línea recta, se admite de forma implícita que el análisis carece de sesgo significativo, que la respuesta es proporcional a la concentración de analito y que la dispersión de los errores aleatorios es constante en todo el intervalo de interés. En la mayoría de los casos, estas hipótesis se basan en la experiencia acumulada durante el desarrollo del método o durante un plazo más largo y son, en consecuencia, razonablemente fiables. No obstante, la práctica correcta de la metrología se basa en hipótesis comprobadas. Por este motivo, numerosos estudios de validación se basan en la comprobación estadística de hipótesis con el fin de realizar una comprobación básica de que las hipótesis razonables planteadas en relación con los principios del método no presentan deficiencias graves.

Esta proposición aparentemente abstrusa presenta una importante consecuencia práctica: es más fácil comprobar si una hipótesis fiable presenta alguna desviación de importancia que “demostrar” que una determinada hipótesis es correcta. De este modo, cuando una determinada técnica analítica (como el análisis mediante cromatografía en fase gaseosa o los métodos de digestión en medio ácido) se ha utilizado con éxito durante mucho tiempo, en diversos analitos y matrices, las comprobaciones para su validación consisten, justificadamente, en ensayos relativamente sencillos, como medida de precaución. Por el contrario, cuando la experiencia acumulada es escasa, el estudio de validación debe proporcionar pruebas concluyentes de que las hipótesis planteadas son válidas en los casos particulares objeto de estudio y generalmente se deberán estudiar todas las posibles circunstancias de forma pormenorizada. Por consiguiente, en cada caso específico, el grado de profundidad de los estudios de validación necesarios dependerá, en parte, de la experiencia acumulada sobre la técnica de analítica en cuestión.

En el razonamiento que se expone a continuación se presupone que el laboratorio tiene suficiente experiencia práctica en la técnica de interés y que la finalidad de cualesquiera pruebas de significación que se realicen es comprobar que no existen pruebas concluyentes que contradigan las hipótesis en las que se basa el protocolo objeto de estudio. El lector deberá tener presente que en el caso de técnicas de medición no habituales o menos arraigadas puede ser necesario realizar comprobaciones más rigurosas.

4.3 Fuentes de error en el análisis

Los errores en las mediciones analíticas proceden de diferentes fuentes* y en diferentes niveles de organización. Una clasificación útil de estas fuentes (para una determinada concentración de analito)** [24] es la siguiente:

- error aleatorio de la medición (repetibilidad)
- efecto del proceso analítico (se manifiesta como sesgo en un proceso simple y como variación aleatoria en varios procesos)
- efecto del laboratorio (se manifiesta como sesgo de un solo laboratorio)
- sesgo del método
- efecto de variación de la matriz

* En el presente documento no se estudia la incertidumbre de muestreo en el sentido estricto de la incertidumbre debida a la preparación de la muestra de laboratorio a partir de la masa objeto de análisis. La incertidumbre asociada a la toma de una porción de ensayo de la muestra de laboratorio es una parte inseparable de la incertidumbre de la medición y se incluye automáticamente en varios niveles en el siguiente análisis.

** Las fuentes de error se pueden agrupar o “dividir” de muchas posibles formas que pueden ser útiles para estudiar fuentes específicas de error de forma más detallada o en diversos tipos de situaciones. Por ejemplo, el modelo estadístico de la Norma ISO 5725 generalmente combina los efectos del laboratorio y del proceso analítico, mientras que el procedimiento de estimación de la incertidumbre de la guía ISO GUM es adecuado para evaluar los efectos de cada influencia independiente y mensurable sobre el resultado.

Aunque estas fuentes diferentes pueden no ser necesariamente independientes, la anterior relación es útil para comprobar en qué medida un determinado estudio de validación aborda las diversas fuentes de error.

El término correspondiente a la repetibilidad (dentro del proceso analítico) comprende aportaciones de errores generados en cualquier parte del procedimiento que varía dentro de un proceso analítico, incluidas las aportaciones de los conocidos errores gravimétricos y volumétricos, la heterogeneidad del material de ensayo y la variación producida en las etapas de tratamiento químico del análisis; se comprueba fácilmente observando la dispersión de análisis repetidos. El efecto del proceso analítico comprende otras variaciones diarias que se producen en el sistema de análisis, como cambios de analista, de lotes de reactivos, recalibración de instrumentos, y en el ambiente del laboratorio, (p. ej. cambios de temperatura). En la validación por un solo laboratorio, el efecto del proceso analítico se estima normalmente realizando un experimento diseñado con análisis repetidos de un material adecuado en varios procesos independientes. La variación interlaboratorios procede de factores como variaciones en los métodos de calibración, diferencias entre interpretaciones de un protocolo por los diferentes laboratorios, diferentes fuentes de los equipos o de los reactivos, o cambios de los factores ambientales, tales como diferencias en las condiciones climáticas medias. La existencia real de la variación interlaboratorios se pone claramente de manifiesto en los resultados de ensayos en colaboración (estudios del funcionamiento de métodos) y de ensayos de aptitud; en ocasiones, los resultados de este último tipo de ensayos permiten discernir la variación entre métodos.

En general, la repetibilidad, el efecto del proceso analítico y el efecto del laboratorio son de magnitudes similares, de manera que ninguno de estos factores puede ignorarse sin consecuencias en la validación. En el pasado, se ha tendido a descuidar determinados aspectos, particularmente los relativos a la estimación y notificación de la información sobre la incertidumbre, lo que da lugar a intervalos de incertidumbre excesivamente pequeños. Por ejemplo, los ensayos en colaboración, cuando se realizan de la forma habitual, no describen la situación de forma completa, porque no se estiman las aportaciones a la incertidumbre que derivan del sesgo del método y de la variación de la matriz y se deben abordar de forma independiente (habitualmente mediante un estudio previo en un solo laboratorio). En la validación por un solo laboratorio existe el riesgo específico de obviar además el sesgo del laboratorio y este componente es normalmente el que más contribuye a la incertidumbre de entre la lista antes indicada. Por consiguiente, en la validación por un solo laboratorio se debe prestar atención de forma específica al sesgo del laboratorio.

Además de los problemas mencionados hasta ahora, la validación de un método se limita a su ámbito de aplicación, es decir, al método en tanto se aplica a una determinada categoría de material de ensayo. Si en la categoría de material definida existe una variación sustancial de tipos de matrices, existirá una fuente de variación adicional debida a efectos de la matriz dentro de su categoría. Naturalmente, si el método se utiliza posteriormente para materiales no incluidos en la categoría definida (es decir, no incluidos en el ámbito de aplicación de la validación), no se puede considerar validado el sistema de análisis; se introduce un error adicional de magnitud desconocida en el proceso de medición.

Es también importante que los analistas tengan en cuenta cómo varía el funcionamiento del método en función de la concentración del analito. En la mayoría de los casos, la dispersión de los resultados aumenta invariablemente con la concentración y la recuperación puede ser sustancialmente distinta a concentraciones bajas y altas. Por consiguiente, la incertidumbre de la medición asociada a los resultados depende frecuentemente de ambos efectos, así como de otros factores dependientes de la concentración. Afortunadamente, a menudo es razonable suponer que existe una relación sencilla entre el funcionamiento y la concentración de analito; lo más habitual es suponer que los errores son proporcionales a la concentración de analito*. No obstante, cuando interesa conocer el funcionamiento del método a concentraciones sustancialmente diferentes, es importante comprobar la hipótesis de la relación entre el funcionamiento y la concentración de analito. Normalmente, se comprueba el funcionamiento en los extremos del intervalo probable o en unos pocos niveles seleccionados. Se puede también obtener información del mismo tipo mediante comprobación de ensayos de la linealidad.

* Esto puede no ser cierto para concentraciones inferiores a 10 veces el límite de detección.

4.4 Efectos del método y del laboratorio

En la validación de métodos por un solo laboratorio es fundamental tener en cuenta el sesgo del método y el efecto del laboratorio. Existen unos pocos laboratorios con instalaciones especiales en los que dichos sesgos se pueden considerar insignificantes, pero esta circunstancia es absolutamente extraordinaria. (No obstante, si un determinado análisis se realiza únicamente en un laboratorio, los sesgos del método y del laboratorio se deben analizar desde un punto de vista diferente.) Normalmente, los efectos del método y del laboratorio deben tenerse en cuenta en la cuantificación de la incertidumbre, pero frecuentemente son más difíciles de abordar que el error asociado a la repetibilidad y el efecto de la serie. En general, para evaluar las respectivas incertidumbres es necesario utilizar información obtenida de forma independiente del laboratorio. Las fuentes generalmente más útiles de esta información son: a) estadísticas de ensayos en colaboración (no disponibles en muchos casos de validación de métodos por un solo laboratorio), b) estadísticas de ensayos de aptitud y c) resultados del análisis de materiales de referencia certificados.

En los ensayos en colaboración se estima directamente la varianza de los sesgos entre laboratorios. Aunque el diseño de estos ensayos puede presentar deficiencias teóricas, estas estimaciones de la varianza son apropiadas para numerosos fines prácticos. En consecuencia, siempre es instructivo comprobar la validación por un solo laboratorio mediante la comparación de las estimaciones de la incertidumbre con las estimaciones de la reproducibilidad obtenidas en ensayos en colaboración. Si el resultado del ensayo en un solo laboratorio es substancialmente menor que el de los ensayos en colaboración, probablemente se han menospreciado importantes fuentes de incertidumbre. (Por el contrario, puede darse el caso de que un determinado laboratorio funcione con una incertidumbre menor que la obtenida en los ensayos en colaboración; en este caso, el laboratorio deberá tomar medidas especiales para justificar dicha afirmación.) Si no se ha realizado ningún ensayo en colaboración de la combinación específica de método y material de ensayo, se puede obtener normalmente una estimación de la desviación típica de la reproducibilidad σ_H a una concentración de analito c superior a unas 120 ppb mediante la función de Horwitz, $\sigma_H = 0,02c^{0,8495}$, donde ambas variables se expresan como fracciones máscas. (El valor estimado de Horwitz normalmente se diferencia en menos de un factor de alrededor de dos de los resultados observados en el estudio en colaboración.) Se ha observado que la función de Horwitz no es correcta en fracciones inferiores a alrededor de 120 ppb y que en esos casos es más adecuado aplicar una función modificada [21,25]. Toda esta información se puede trasladar, con cambios mínimos, al ámbito de los estudios en un solo laboratorio.

Son particularmente interesantes las estadísticas de los ensayos de aptitud porque proporcionan información general sobre la magnitud de los sesgos del laboratorio y del método combinado y proporcionan al participante información sobre el error total en determinadas ocasiones. Estadísticas como la desviación típica robusta de los resultados de los laboratorios participantes, para un analito, en una ronda del ensayo, se pueden utilizar, en principio, de forma similar a las desviaciones típicas de la reproducibilidad de los ensayos en colaboración, es decir, para obtener una referencia de la incertidumbre global para su comparación con estimaciones individuales de validaciones por un solo laboratorio. En la práctica, puede ser más difícil acceder a las estadísticas de los ensayos de aptitud, porque no se tabulan y publican de forma sistemática, como las de los ensayos en colaboración, sino que únicamente se ponen a disposición de los participantes. Por supuesto, para que estas estadísticas tengan utilidad deben referirse a la matriz y concentración de analito pertinentes. Cada uno de los participantes en programas de ensayos de aptitud puede también evaluar la validez de la incertidumbre estimada comparando sus resultados comunicados con los valores asignados de rondas sucesivas [26]. Se trata, sin embargo, de una actividad en continuo progreso y, por consiguiente, no se incluye, en rigor, en el alcance de la validación por un solo laboratorio (una actividad que se realiza una sola vez).

Si se dispone de material de referencia certificado (MRC) adecuado, se puede determinar en un ensayo por un solo laboratorio, mediante el análisis del MRC en repetidas ocasiones, el sesgo del laboratorio y el sesgo del método combinados. El valor estimado del sesgo combinado es la diferencia entre el resultado medio y el valor certificado.

No siempre se dispone de MRC adecuados, por lo que puede ser obligado utilizar otros materiales. En ocasiones, se pueden utilizar para este fin materiales sobrantes de los ensayos de aptitud y, aunque los valores asignados a los materiales pueden entrañar incertidumbres cuestionables, pueden utilizarse sin duda para comprobar el sesgo total. Concretamente, los valores asignados de ensayos de aptitud se eligen generalmente de

forma que proporcionen una estimación con un sesgo mínimo, de suerte que es aconsejable comprobar si existe un sesgo significativo con dicho material. Otra posibilidad es obtener estimaciones de estos sesgos a partir de información sobre adición (*spiking*) y recuperación [4], aunque estas técnicas pueden llevar asociadas fuentes de incertidumbre no mensurables.

Actualmente, el efecto menos reconocido en la validación es el debido a la variación de las matrices dentro de la categoría de material de ensayo definida. El requisito teórico para la estimación de este componente de incertidumbre es analizar en una sola serie de un conjunto representativo de materiales de ensayo, estimar el sesgo de cada uno de los materiales y calcular la varianza de estos sesgos. (El análisis en una sola serie implica que los sesgos de nivel superior no tienen efecto alguno en la varianza. Si el intervalo de concentración es amplio, se deben tener en cuenta los cambios del sesgo en función de la concentración.) Si los materiales representativos son MRC, los sesgos se pueden estimar directamente como las diferencias entre los resultados y los valores de referencia, lo que simplifica el procedimiento. En el caso más probable de que no se disponga de suficientes MRC, se puede recurrir, con la debida precaución, a ensayos de recuperación con una gama de materiales de ensayo típicos. Actualmente, se dispone de muy poca información cuantitativa sobre la magnitud de las incertidumbres de esta fuente, aunque en algunos casos se sospecha que son grandes.

5. REALIZACIÓN DE ESTUDIOS DE VALIDACIÓN

El diseño y ejecución pormenorizados de estudios de validación de métodos se describe ampliamente en otros lugares y no se repetirá en el presente documento. No obstante, se consideran relevantes los principios fundamentales que se plantean a continuación.

Es indispensable que los estudios de validación sean representativos. Es decir, los estudios se deben realizar, en la medida de lo posible, de forma que proporcionen una muestra realista del número y variedad de efectos que se producen durante la aplicación normal del método y de forma que abarquen los intervalos de concentración y tipos de muestras correspondientes al alcance del método. Cuando un factor (como la temperatura ambiente) ha variado de forma representativa y aleatoria durante la realización de un experimento de precisión, por ejemplo, los efectos de ese factor se manifiestan de forma directa en la varianza observada y no se necesita ningún estudio adicional, a menos que se desee una optimización adicional del método.

En el contexto de la validación de métodos, se considera que un factor tiene una “variación representativa” cuando presenta una distribución de valores conforme al intervalo previsto del parámetro en cuestión. En el caso de parámetros mensurables continuos, se puede tratar de un intervalo permitido, una incertidumbre declarada o un intervalo previsto; para factores discontinuos o factores con efectos imprevisibles como la matriz de muestra, un intervalo representativo corresponde a la variedad de tipos o “niveles de los factores” permitidos o existentes en la aplicación normal del método. Teóricamente, la representatividad se extiende no sólo al intervalo de valores, sino también a su distribución. Desafortunadamente, con frecuencia no es rentable disponer todas las variaciones posibles de numerosos factores a numerosos niveles. No obstante, para la mayoría de los fines prácticos, los ensayos basados en los extremos del intervalo previsto, o en cambios mayores que los previstos, constituyen un mínimo aceptable.

Al seleccionar factores en relación con la variación, es importante asegurar que se “apliquen” los efectos mayores en la mayor medida posible. Por ejemplo, cuando la variación día tras día (quizá debida a efectos de la recalibración) es substancial comparada con la repetibilidad, dos determinaciones en cinco días diferentes proporcionarán una mejor estimación de la precisión intermedia que cinco determinaciones en dos días diferentes. Aún mejor será la realización de diez determinaciones únicas en días diferentes, sometidas a un control suficiente, aunque esta estrategia no proporcionará información adicional sobre la repetibilidad interdiaria.

Es evidente que en la planificación de las pruebas de significación, los estudios deben tener suficiente capacidad estadística para detectar estos efectos antes de que adquieran importancia práctica (es decir, de que sean de magnitud semejante al componente mayor de la incertidumbre).

Además, pueden ser importantes las siguientes consideraciones:

- Cuando se conoce o se sospecha la interacción de factores, es importante asegurarse de que se tiene en cuenta el efecto de la interacción. Esto se puede conseguir asegurando la selección aleatoria de diferentes niveles de los parámetros objeto de la interacción, o bien mediante un diseño sistemático cuidadoso para obtener los efectos de la “interacción” o información sobre la covarianza.
- Al realizar estudios sobre el sesgo total, es importante que los materiales de referencia y los valores correspondan a los materiales sometidos a ensayos sistemáticos.

6. ALCANCE DE LOS ESTUDIOS DE VALIDACIÓN

La medida en que un laboratorio debe emprender la validación de un método nuevo, modificado o poco conocido depende hasta cierto punto de la condición efectiva del método y de la capacidad del laboratorio. A continuación se dan indicaciones sobre el grado de las medidas de validación y verificación para diferentes circunstancias. Excepto cuando se indica lo contrario, se entiende que la finalidad del método es el análisis sistemático.

- *El laboratorio deberá utilizar un método validado “plenamente”.* El método ha sido estudiado en un ensayo en colaboración, por lo que el laboratorio debe comprobar que es capaz de lograr las características de funcionamiento del método publicadas (o que es capaz de cumplir de otro modo los requisitos de la labor de análisis). El laboratorio deberá realizar estudios de precisión, estudios de sesgo (incluidos estudios de variación de la matriz) y posiblemente estudios de linealidad, aunque se pueden omitir algunos ensayos, como los de robustez.
- *El laboratorio deberá utilizar un método validado plenamente, pero deberá utilizarse una nueva matriz.* El método ha sido estudiado en un ensayo en colaboración, por lo que el laboratorio deberá comprobar que la nueva matriz no introduce nuevas fuentes de error en el sistema. Deberán realizarse los mismos estudios de validación que en el caso anterior.
- *El laboratorio deberá utilizar un método bien establecido, pero que no se ha estudiado en colaboración.* Deberán realizarse los mismos estudios de validación que en el caso anterior.
- *El método se ha publicado en publicaciones científicas junto con algunas características analíticas.* El laboratorio deberá realizar estudios de precisión, de sesgo (incluidos estudios de variación de la matriz), de robustez y de linealidad.
- *El método se ha publicado en publicaciones científicas, pero sin indicación de las características, o se ha desarrollado de forma interna.* El laboratorio deberá realizar estudios de precisión, de sesgo (incluidos estudios de variación de la matriz), de robustez y de linealidad.
- *El método es empírico.* Un método empírico es aquel en que la magnitud estimada es simplemente el resultado que se obtiene siguiendo el procedimiento indicado. Se diferencia de las mediciones realizadas para evaluar magnitudes independientes del método, como la concentración de un determinado analito en una muestra, en que el sesgo del método es convencionalmente cero y en que la variación de la matriz (es decir, la variación en la categoría definida) es insignificante. No se puede desestimar el sesgo del laboratorio, pero es probablemente difícil estimarlo mediante un experimento en un solo laboratorio. Además, no es probable que se disponga de materiales de referencia. En ausencia de datos de ensayos en colaboración, se puede obtener una cierta estimación de la precisión interlaboratorios mediante un estudio de robustez diseñado con ese objetivo, o se puede estimar mediante la función de Horwitz.
- *Se trata de un análisis especial.* Ocasionalmente, se necesita realizar un análisis especial para determinar, sin un gasto elevado y con escasa criticidad, el margen general de un valor. En consecuencia, la inversión que se puede dedicar a la validación es muy limitada. El sesgo se debería estudiar mediante métodos como la estimación de la recuperación o las adiciones de analito, y la precisión mediante el replicado.
- *Cambios en el personal y en los equipos de laboratorio.* Son ejemplos importantes los siguientes: cambios en los instrumentos principales; lotes nuevos de reactivos muy variables (por ejemplo,

anticuerpos policlonales); cambios realizados en el recinto en el que se ubica el laboratorio; métodos utilizados por primera vez por empleados nuevos o el uso de un método validado tras un período sin utilizarlo. En estos casos, el objetivo primordial es demostrar que no se han producido cambios deletéreos. La comprobación mínima es un solo ensayo de comprobación del sesgo; un experimento basado en las situaciones “anterior y posterior” realizado con materiales de ensayo típicos o materiales de control. En general, los ensayos realizados deben reflejar el posible efecto del cambio en el procedimiento de análisis.

7. RECOMENDACIONES

Se formulan las siguientes recomendaciones relativas a la utilización de la validación de métodos por un solo laboratorio:

- Cuando sea posible y práctico, los laboratorios deberán utilizar métodos de análisis cuyas características de funcionamiento han sido evaluadas mediante ensayos en colaboración ajustados a un protocolo internacional.
- Cuando no se dispongan de dichos métodos, antes de utilizar un método para recabar datos analíticos para un cliente, se debe validar internamente.
- En la validación por un solo laboratorio, éste deberá seleccionar las características apropiadas para la evaluación de entre las siguientes: aplicabilidad, selectividad, calibración, exactitud, precisión, intervalo, límite de cuantificación, límite de detección, sensibilidad y robustez. Al elegir las características que habrán de determinarse, el laboratorio deberá tener en cuenta las necesidades del cliente.
- Si el cliente lo solicita, el laboratorio deberá proporcionar los datos que demuestran que ha evaluado estas características.

APÉNDICE A: NOTAS SOBRE LOS REQUISITOS PARA EL ESTUDIO DE LAS CARACTERÍSTICAS DE FUNCIONAMIENTO DE LOS MÉTODOS

Los requisitos generales de cada una de las características de funcionamiento de un método son las siguientes.

A1. Aplicabilidad

La documentación que se elabora tras la validación deberá proporcionar, además de las especificaciones de funcionamiento, la siguiente información:

- identidad del analito, incluida la especiación cuando sea necesario (por ejemplo, “arsénico total”);
- intervalo de concentración que abarca la validación (por ejemplo, “0 a 50 ppm”);
- especificación de la gama de matrices del material de ensayo que cubre la validación (por ejemplo, “alimentos marinos”);
- un protocolo con la descripción de los equipos, reactivos, procedimiento (con indicación del intervalo de variación admisible de los datos especificados en las instrucciones, por ejemplo: “calentar a 100 ± 5 °C durante 30 ± 5 min”), procedimientos de calibración y de control de la calidad y las medidas especiales de seguridad que fueran necesarias;
- la aplicación a que se destina y sus requisitos de incertidumbre críticas (por ejemplo, “Análisis de alimentos para fines de clasificación. La incertidumbre típica $u(c)$ del resultado c debe ser menor que $0,1 \times c$ ”).

A2. Selectividad

La selectividad es el grado de exactitud que puede alcanzar un método en la cuantificación del analito en presencia de interferentes. Lo ideal sería evaluar la selectividad con respecto a cualquier interferente importante cuya presencia sea probable. Es particularmente importante comprobar los interferentes que probablemente (basándose en los principios químicos) responden al ensayo. Por ejemplo, es razonable esperar que los ensayos colorimétricos de amoníaco respondan a las aminas alifáticas primarias. Puede que no sea posible, en la práctica, tener en cuenta o ensayar todos los posibles interferentes; en este caso, se recomienda comprobar los casos

probablemente más desfavorables. Por regla general, la selectividad debe ser suficiente para ignorar las posibles interferencias.

En muchos tipos de análisis, la selectividad es esencialmente una evaluación cualitativa basada en si los ensayos de interferencia adecuados producen o no un resultado significativo. Existen, sin embargo, mediciones cuantitativas de la interferencia útiles; por ejemplo, el índice de selectividad b_{an}/b_{int} , en que b_{an} es la sensibilidad del método (en decir, la pendiente de la función de calibración) y b_{int} la pendiente de la respuesta producida de forma independiente por un posible interferente. El valor de b_{int} se puede determinar de forma aproximada aplicando el procedimiento con un blanco de la matriz y el mismo blanco al que se ha añadido una concentración adecuada del posible interferente. Si no se dispone de un blanco de la matriz y se utiliza en su lugar un material típico, se puede calcular b_{int} mediante este experimento sencillo sólo si se acepta la hipótesis de que no existen efectos mutuos de la matriz. Obsérvese que b_{int} se determina más fácilmente en ausencia del analito, porque cuando el interferente afecta a la sensibilidad del propio analito (un efecto de la matriz), el efecto se puede confundir con otro tipo de interferencia.

A3. Calibración y linealidad

Exceptuando errores graves que pudieran producirse en la preparación de los materiales de calibración, los errores de calibración representan habitualmente (pero no siempre) un componente menor de la incertidumbre total y habitualmente se pueden incorporar sin riesgo a diversas categorías estimadas mediante métodos que van de lo general a lo particular (“*top-down*”). Por ejemplo, los errores aleatorios producto de la calibración son parte del sesgo de la serie, que se evalúa de forma global, mientras que los errores sistemáticos del mismo tipo pueden aparecer como parte del sesgo del laboratorio, que también se evalúa de forma global. No obstante, es útil conocer algunas de las características de la calibración al comienzo de la validación de un método porque afectan a la estrategia para el desarrollo óptimo del procedimiento; por ejemplo, si la función de calibración a) es lineal, b) pasa por el origen y c) no se ve afectada por la matriz del material de ensayo. Los procedimientos que se describen aquí se refieren a estudios relativos a la calibración en la validación, que son por necesidad más exigentes que los relativos a la calibración realizada durante los análisis sistemáticos. Por ejemplo, una vez que se ha determinado en la validación que una función de calibración es lineal y pasa por el origen, se puede utilizar para el análisis sistemático una estrategia de calibración mucho más sencilla (por ejemplo, un diseño repetido de dos puntos). Los errores de esta estrategia de calibración más sencilla se incluirán normalmente, para los fines de la validación, en las fuentes de error de nivel más elevado.

A3.1 Linealidad e intercepción

La linealidad se puede comprobar de manera práctica examinando una representación gráfica de los residuos obtenidos de la regresión lineal de las respuestas a las concentraciones en un conjunto de muestras de calibración adecuadas. Una línea curva indica una posible falta de ajuste debido a que la función de calibración no es lineal. Se puede realizar una prueba de significación comparando la varianza de falta de ajuste con la varianza debida al error puro. No obstante, en determinados tipos de calibración analítica pueden producirse causas de falta de ajuste diferentes de la no linealidad, por lo que a la prueba de significación debe acompañar un gráfico de los residuos. A pesar de que actualmente se utiliza de forma generalizada el coeficiente de correlación como indicador de la calidad del ajuste, su uso para comprobar la linealidad es engañoso e inadecuado y no debe utilizarse para este fin.

El diseño es fundamental en las pruebas de falta de ajuste, porque es fácil confundir la falta de linealidad con la deriva. Si no existe una estimación independiente del error puro, se debe estimar mediante mediciones repetidas. A falta de directrices específicas, se deben aplicar las siguientes (para la calibración lineal de una variable):

- se debe disponer de al menos seis patrones de calibración;
- los patrones de calibración se deben distribuir de forma regular en el intervalo de concentraciones de interés;
- el intervalo debe abarcar del 0 al 150 % o del 50 al 150 % de la concentración probable, según cuál de los dos intervalos sea más adecuado y
- los patrones de calibración se deben analizar al menos por duplicado y preferentemente por triplicado o más veces, en un orden aleatorio.

Tras un ajuste exploratorio mediante una regresión lineal simple, se debe examinar si existen tendencias evidentes en los valores de los residuos. En la calibración analítica, es bastante frecuente la heteroscedasticidad, y si se observa una tendencia que sugiera su existencia, es preferible procesar los datos de la calibración mediante una regresión ponderada. Si en estas circunstancias no se aplica una regresión ponderada, se producirían errores exagerados en el extremo inferior de la función de calibración.

La prueba de falta de ajuste se puede realizar mediante una regresión simple o ponderada. Si no hay una falta de ajuste significativa, también se pueden someter los datos a una prueba para determinar si la intercepción es significativamente distinta de cero.

A3.2 *Ensayo del efecto general de la matriz*

Si los patrones de calibración se pueden elaborar como simples soluciones del analito, la calibración se simplifica enormemente. Si se adopta esta estrategia, se deben evaluar en la validación los efectos de un posible mal apareamiento general de las matrices. Se puede realizar un ensayo del efecto general de la matriz aplicando el método de adiciones de analito (también llamado de “adiciones estándar”) a una solución de ensayo derivada de un material de ensayo típico. El ensayo debería realizarse de forma tal que se obtenga la misma dilución final que con el procedimiento normal y las adiciones deberían abarcar el mismo intervalo que la validación de la calibración definida por el procedimiento. Si la calibración es lineal, se pueden comparar las pendientes de la función de calibración habitual con la representación gráfica de las adiciones de analito para determinar si existe una diferencia significativa. Si la diferencia no es significativa, no existe un efecto general de la matriz detectable. Si la calibración no es lineal, la prueba de significación deberá basarse en un método más complejo, pero habitualmente basta una comparación visual a concentraciones iguales. Si el resultado de esta prueba no es significativo, indicará a menudo que tampoco existe un efecto de variación de la matriz (sección A13).

A3.3 *Procedimiento de calibración final*

Puede también ser necesario validar de forma independiente la estrategia de calibración especificada en el procedimiento, aunque los errores que ello comporta contribuirán a la incertidumbre estimada de forma conjunta. Lo importante en este sentido es que la incertidumbre estimada a partir de los diseños específicos para la linealidad, etc., serán menores que las incertidumbres derivadas de la calibración más sencilla definida en el protocolo del procedimiento.

A4. Veracidad

A4.1 *Estimación de la veracidad*

La veracidad es la proximidad entre el resultado de un ensayo y el valor de referencia aceptado de la propiedad objeto de medición. La veracidad se expresa en términos cuantitativos como “sesgo”; cuanto menor es el sesgo, mayor es la veracidad. Típicamente, el sesgo se determina comparando la respuesta obtenida aplicando el método a un material de referencia con el valor asignado conocido del material. Se recomienda realizar una prueba de significación. Cuando la incertidumbre del valor de referencia no es insignificante, la evaluación de los resultados debe tener en cuenta dicha incertidumbre además de la variabilidad estadística.

A4.2 *Condiciones para la realización de experimentos de estimación de la veracidad*

En un sistema de análisis, el sesgo puede manifestarse en diferentes niveles, por ejemplo, sesgo de la serie, sesgo del laboratorio y sesgo del método. Es importante recordar a cuál de éstos corresponde cada uno de los diversos métodos de estimación del sesgo. Concretamente:

- La media de un conjunto de análisis de un material de referencia, realizados todos en una sola serie, proporciona información sobre la suma de los efectos del método, del laboratorio y de la serie de esa serie concreta. Como se supone que el efecto interserial es aleatorio, la variación interserial de los resultados será mayor que la variación previsible a tenor de la dispersión observable de los resultados, lo cual se debe tener en cuenta en la evaluación de los resultados (por ejemplo, comprobando el sesgo medido con la desviación típica interserial determinada de forma independiente).

- La media de análisis repetidos de un material de referencia en varias series es una estimación del efecto combinado de los sesgos del método y del laboratorio en el laboratorio específico (excepto cuando el valor se asigna utilizando el método específico).

A4.3 Valores de referencia para experimentos de estimación de la veracidad

A4.3.1 Materiales de referencia certificados (MRC)

Los MRC se pueden relacionar con patrones internacionales con una incertidumbre conocida y se pueden utilizar por consiguiente para evaluar simultáneamente todos los componentes del sesgo (del método, interlaboratorios e intralaboratorio), suponiendo que no existe desigualdad de las matrices. En consecuencia, en la validación de la veracidad se deben utilizar MRC cuando sea factible. Es importante asegurarse de que las incertidumbres de los valores certificados son suficientemente pequeñas para permitir la detección de sesgos cuya magnitud sea importante. Cuando no lo sean, se recomienda en cualquier caso la utilización de MRC, pero se deberán realizar comprobaciones adicionales.

En un experimento de estimación de la veracidad típico se genera una respuesta media correspondiente a un material de referencia. En la interpretación del resultado, se debe tener en cuenta la incertidumbre asociada al valor certificado, junto con la incertidumbre debida a la variación estadística en el laboratorio. Según el objetivo del experimento, este último término se puede basar en la desviación típica intraserial o interserial, o en una estimación de la desviación típica interlaboratorios. Cuando la incertidumbre del valor certificado es pequeña, normalmente se realiza una prueba *t* de Student, con el término de precisión adecuado.

En caso necesario y factible, se deben examinar varios MRC convenientes, con matrices y concentraciones de analito adecuadas. Cuando se hace así y las incertidumbres de los valores certificados son menores que las de los resultados analíticos, es razonablemente seguro evaluar los resultados mediante regresión simple. De esta forma, el sesgo se expresaría en función de la concentración y aparecería como ordenada en el origen distinta de cero (sesgo “transicional” o constante) o como pendiente no unitaria (sesgo “rotacional” o proporcional). Cuando existe una amplia gama de matrices, los resultados se deben interpretar con la precaución debida.

A4.3.2 Materiales de referencia

Cuando no se dispone de MRC, o además de los MRC, se puede utilizar cualquier material suficientemente bien caracterizado para este fin (un material de referencia [10]), teniendo siempre presente que si bien un sesgo no significativo no demuestra que el sesgo es cero, un sesgo significativo en cualquier material es motivo para hacer una investigación. Son ejemplos de materiales de referencia los materiales caracterizados por un productor de materiales de referencia pero con valores que no van acompañados de una declaración de la incertidumbre u otra forma de caracterización; materiales caracterizados por un fabricante del material; materiales caracterizados en el laboratorio para su utilización como materiales de referencia y materiales sometidos a una aplicación *round-robin* restringida o distribuidos en un ensayo de aptitud. Si bien la rastreabilidad de estos materiales puede ser cuestionable, es mucho mejor utilizarlos que renunciar a evaluar el sesgo. Los materiales se utilizarían de forma muy similar a los MRC aunque, al no declararse la incertidumbre, cualquier prueba de significación se debe basar enteramente en la precisión observable de los resultados.

A4.3.3 Utilización de un método de referencia

En principio, se puede utilizar un método de referencia para comprobar el sesgo en otro método objeto de validación. Esta es una posibilidad útil para comprobar la validez de otro método o de una modificación de un método convencional establecido, ya validado y utilizado en el laboratorio. Se analizan varios materiales de ensayo típicos con ambos métodos, preferiblemente abarcando de forma bastante uniforme un intervalo de concentración útil. La comparación de los resultados correspondientes a las diversas concentraciones del intervalo mediante un método estadístico adecuado (por ejemplo, una prueba *t* de Student para datos apareados, con las comprobaciones debidas de la homogeneidad de la varianza y de la normalidad) permitiría comprobar la posible existencia de un sesgo entre los métodos.

A4.3.4 Método de la Adición (*spiking*) y recuperación

Si no se dispone de materiales de referencia, o para respaldar los resultados de estudios de materiales de referencia, se puede estimar el sesgo mediante el método de adición y recuperación. Se analiza un material de ensayo típico mediante el método objeto de validación, tanto en su estado original como tras la adición (*spiking*) de una masa conocida del analito a la porción de ensayo. La diferencia entre los dos resultados expresada como proporción de la masa añadida se conoce como recuperación de sustitución o, a veces, recuperación marginal. Las recuperaciones sensiblemente diferentes de la unidad indican la existencia de un sesgo que afecta al método. En rigor, los estudios de recuperación según se describen aquí únicamente evalúan el sesgo debido a los efectos sobre el analito añadido; estos efectos no necesariamente afectan de la misma forma al analito nativo al que además pueden afectar otros efectos diferentes. En consecuencia, en los estudios de adición y recuperación se cumple claramente la siguiente observación: si bien una buena recuperación no es garantía de veracidad, una mala recuperación es ciertamente una indicación de falta de veracidad. Se han descrito de forma pormenorizada en otras fuentes [4] métodos de tratamiento de los datos de adición y recuperación.

A5. Precisión

La precisión es la estrecha conformidad entre resultados de ensayos independientes obtenidos en condiciones estipuladas. Se expresa habitualmente como desviación típica o como desviación típica relativa. La distinción entre precisión y sesgo es fundamental, pero depende del nivel en el que se contempla el sistema de análisis. Así, desde el punto de vista de una sola determinación, cualquier desviación que afecte a la calibración de la serie se puede considerar un sesgo. Desde el punto de vista del analista que revisa el trabajo de un año, el sesgo del proceso analítico será diferente cada día y actuará como variable aleatoria con una precisión asociada. Las condiciones estipuladas para la estimación de la precisión tienen en cuenta ambos puntos de vista.

Para la validación por un solo laboratorio, se deben tener en cuenta dos tipos de condiciones: a) precisión en condiciones de repetibilidad, que describe las variaciones observadas durante un solo proceso analítico como esperanza θ y desviación típica σ_r , y b) precisión en condiciones de proceso a proceso analítico, que describe las variaciones del sesgo de las series δ_{run} como esperanza θ , desviación típica σ_{run} . Habitualmente, los resultados analíticos individuales se ven afectados por ambas fuentes de error, cuya precisión combinada es, en consecuencia, $\sigma_{tot} = (\sigma_r^2/n + \sigma_{run}^2)^{1/2}$, en que n es el número de resultados repetidos promediados en un proceso analítico para obtener el resultado comunicado. La forma más sencilla de obtener las dos estimaciones de la precisión es analizar el material de ensayo seleccionado por duplicado en varios procesos analíticos sucesivos. Se pueden entonces calcular los diferentes componentes de la varianza mediante un análisis de varianza unifactorial. Cada análisis duplicado debe consistir en una ejecución independiente del procedimiento aplicado a porciones de ensayo diferentes. De otro modo, la precisión combinada σ_{tot} se puede estimar directamente analizando el material de ensayo una sola vez en procesos analíticos sucesivos y estimando la desviación típica mediante la ecuación habitual. (Obsérvese que las desviaciones típicas muestrales se identifican generalmente mediante el símbolo s , para distinguirlas de las desviaciones típicas poblacionales σ .)

Es importante que los valores de precisión sean representativos de las condiciones de ensayo probables. En primer lugar, la variación de las condiciones entre las series debe ser representativa de lo que ocurriría normalmente en el laboratorio en la aplicación sistemática del método. Por ejemplo, las variaciones de los lotes de reactivos, analistas e instrumentos deben ser representativas. En segundo lugar, la matriz y (preferiblemente) el tamaño de partículas del material de ensayo utilizado debe ser típico de los materiales que se encontrarán probablemente en la aplicación sistemática. De manera que serían adecuados los propios materiales de ensayo o, en menor medida, materiales de referencia con la misma matriz, pero no las soluciones patrón del analito. Obsérvese también que los MRC y los materiales de referencia preparados frecuentemente están más homogeneizados que los materiales de ensayo típicos y la precisión que se obtenga mediante su análisis puede, por consiguiente, subestimar la variación que se observará con los materiales de ensayo.

Muy frecuentemente, la precisión varía en función de la concentración del analito. Habitualmente se aceptan las siguientes hipótesis: i) que la precisión no cambia en función de la concentración del analito o ii) que la desviación típica es proporcional a la concentración del analito o es linealmente dependiente de la misma. Ambas hipótesis deben comprobarse si se espera que la concentración del analito varíe de forma substancial (es decir, en más de alrededor de un 30 % con respecto a su valor central). El experimento menos costoso es

probablemente una simple evaluación de la precisión en o cerca de los extremos del intervalo operativo, junto con una prueba estadística adecuada de la diferencia de las varianzas. La prueba F es adecuada si la distribución de los errores es normal.

Pueden obtenerse datos de precisión de una gran variedad de tipos de condiciones diferentes además de los mínimos de condiciones de repetibilidad y entre procesos analíticos indicados aquí, y puede ser necesario obtener información adicional. Por ejemplo, puede ser útil para la evaluación de los resultados o para mejorar la medición, disponer de estimaciones independientes de los efectos del operador y de proceso analítico, de los efectos interdiarios o intradiarios o de la precisión que se puede alcanzar utilizando un instrumento o varios. Se dispone de diversos diseños y técnicas de análisis estadístico diferentes y es altamente recomendable cuidar el diseño experimental en todos los estudios de este tipo.

A6. Recuperación

Los métodos de estimación de la recuperación se han descrito en un apartado anterior junto con los métodos de estimación de la veracidad.

A7. Intervalo

El intervalo validado es el intervalo de concentración de analito en el cual el método se puede considerar validado. Es importante comprender que este intervalo no es necesariamente idéntico al intervalo útil de la calibración. La calibración puede abarcar un intervalo de concentraciones amplio, pero el resto de la validación (habitualmente una parte mucho más importante en términos de la incertidumbre) abarcará un intervalo más reducido. En la práctica, la mayoría de los métodos se validan a sólo uno o dos niveles de concentración. El intervalo validado se puede establecer mediante una extrapolación razonable de estos puntos en la escala de concentraciones.

Cuando la aplicación del método se centra en una concentración de interés muy superior al límite de detección, se debería validar el método en un valor cercano a ese nivel crítico puntual. Es imposible definir una regla general para la extrapolación segura de este resultado a otras concentraciones de analito porque depende en gran medida del sistema de análisis concreto. Por lo tanto, se debe indicar en el informe del estudio de validación el intervalo en torno al valor crítico en que la persona que realiza la validación, basándose en su criterio profesional, considera que la incertidumbre estimada es válida.

Cuando el intervalo de concentración de interés se aproxima a cero, o al límite de detección, no es correcto suponer que la incertidumbre absoluta es constante ni que lo es la incertidumbre relativa. En esta situación frecuente, una aproximación útil es suponer una relación funcional, con intercepción positiva, entre la incertidumbre u y la concentración c de tipo lineal, es decir, de la forma

$$u(c) = u_0 + \theta c$$

en que θ es la incertidumbre relativa estimada a cierta concentración muy superior al límite de detección; u_0 es la incertidumbre típica estimada para una concentración cero y en algunas circunstancias se puede estimar como $c_L/3$. En estas circunstancias, es razonable considerar que el intervalo validado abarca desde cero a un múltiplo entero pequeño del nivel de validación superior. Nuevamente, dependerá del criterio profesional del analista.

A8. Límite de detección

En sentido general, el límite de detección es la menor cantidad o concentración de analito presente en la muestra de ensayo que se puede distinguir de cero de forma fiable [22,23]. En sistemas de análisis en los que el intervalo de validación no incluye ni se aproxima al límite de detección, no es necesario incluir dicho límite en la validación.

A pesar de la aparente simplicidad del concepto, existen numerosos problemas relacionados con la cuestión del límite de detección, los cuales se resumen a continuación.

- La cuestión se puede abordar desde varios enfoques teóricos posibles, en cada uno de los cuales el límite se define de forma algo diferente. Los intentos de aclarar la cuestión parecen aún más confusos.

- Aunque todos estos enfoques dependen de una estimación de la precisión a una concentración cero o próxima a cero, no está claro si se debe considerar que esto implica condiciones de repetibilidad o alguna otra condición de la estimación.
- A menos que se recoja una cantidad desmesurada de datos, las estimaciones del límite de detección estarán sujetas a una variación aleatoria bastante grande.
- A menudo, las estimaciones del límite de detección están sesgadas a la baja debido a factores operativos.
- Las inferencias estadísticas relativas al límite de detección se basan en la hipótesis de la normalidad que, a concentraciones bajas, es, como mínimo, cuestionable.

Para la mayoría de los fines prácticos de la validación de métodos, al parecer es mejor optar por una definición sencilla que conduzca a una estimación rápida que se utiliza únicamente como orientación aproximada sobre la utilidad del método. Se debe, sin embargo, tener en cuenta que el límite de detección según se estima en el desarrollo del método, puede no ser idéntico en concepto o valor numérico al utilizado para caracterizar un método de análisis completo. Por ejemplo, el “límite de detección instrumental” mencionado en publicaciones científicas o en folletos que acompañan a los instrumentos y ajustado después en función de la dilución, es a menudo mucho menor que el límite de detección “práctico” y no es adecuado para la validación de métodos.

Por consiguiente, para la validación de métodos se recomienda utilizar una estimación de la precisión (S_0) basada en al menos seis determinaciones completas independientes de la concentración de analito en un blanco de una matriz típica o en un material de baja concentración, sin censurar los resultados iguales a cero o negativos, y calcular el límite de detección aproximado como $3S_0$. Obsérvese que con los grados de libertad mínimos recomendados este valor presenta una incertidumbre considerable y bien pudiera ser erróneo por un factor de 2. Cuando se necesitan estimaciones más rigurosas (por ejemplo, para respaldar decisiones basadas en la detección o no de una sustancia, se deben hacer referencia a directrices adecuadas (véanse, por ejemplo, las referencias 22 y 23).

A9. Límite de determinación o límite de cuantificación

En ocasiones, resulta útil establecer una concentración mínima por debajo de la cual se considera que la precisión del método de análisis no es aceptable. Algunas veces esta precisión se define, de forma arbitraria, como el 10 % de la desviación típica relativa; otras veces el límite se considera, de forma igualmente arbitraria, como un múltiplo fijo (típicamente 2) del límite de detección. Aunque resulta en cierto modo tranquilizante trabajar en niveles superiores a dicho límite, se debe reconocer que se trata de una división bastante arbitraria de la escala de concentración; las mediciones en niveles inferiores a dicho límite no carecen de información y pueden ser adecuadas para el fin perseguido. Por lo tanto, en este documento no se recomienda la utilización de este tipo de límite en la validación. Es preferible tratar de expresar la incertidumbre de la medición en función de la concentración y comparar esta función con un criterio de aptitud para los fines acordados entre el laboratorio y el cliente o usuario final de los datos.

A10. Sensibilidad

La sensibilidad de un método es el gradiente de la función de calibración. Como es habitualmente arbitraria y depende de los ajustes instrumentales fijados, no es útil en la validación. (No obstante, puede ser útil en procedimientos de garantía de la calidad, para comprobar si el funcionamiento de un instrumento es constante y satisfactorio.)

A11. Robustez

La robustez de un método de análisis es la resistencia al cambio de los resultados obtenidos mediante un método de análisis cuando se realizan pequeñas modificaciones de las condiciones experimentales descritas en el procedimiento. En el protocolo del método se deben formular los límites de los parámetros experimentales (aunque no siempre se ha hecho así en el pasado), y estas desviaciones admisibles no deben producir, por separado o combinadas, ningún cambio significativo en los resultados obtenidos. (En este contexto, se entiende por “cambio significativo” aquél que haría que el método no se pudiera aplicar respetando los límites de incertidumbre acordados que definen la aptitud para los fines.) Se debe identificar qué aspectos del método

pueden afectar probablemente a los resultados y se debe evaluar su influencia sobre el funcionamiento del método mediante pruebas de robustez.

La robustez de un método se comprueba introduciendo de forma intencionada pequeños cambios en el procedimiento y examinando el efecto de estos cambios en los resultados. Puede ser necesario tener en cuenta diversos parámetros del método, pero como el efecto de la mayoría de éstos será insignificante, normalmente se podrán modificar varios de una vez. Youden [13] ha descrito un experimento de bajo costo basado en diseños factoriales fraccionarios. Por ejemplo, es posible formular un enfoque basado en ocho combinaciones de siete factores variables; es decir, estudiar los efectos de siete parámetros con sólo ocho resultados analíticos. También se pueden aplicar enfoques en los que se modifica una sola variable cada vez.

Algunos de los factores que pueden ser objeto de ensayo de robustez son: cambios en los instrumentos, el analista o la marca de reactivo; concentración de un reactivo; pH de una solución; temperatura de una reacción; tiempo que se deja transcurrir antes de dar por terminado un proceso, etc.

A12. Aptitud para los fines

La aptitud para los fines es el grado de correspondencia entre el funcionamiento de un método y los criterios, acordados entre el analista y el usuario final de los datos, que describen las necesidades del usuario final. Por ejemplo, la magnitud de los errores en los datos no debe ser tal que dé lugar a decisiones equivocadas con una frecuencia mayor que la definida por una probabilidad pequeña establecida, pero los errores no deben ser tan pequeños que supongan un gasto innecesario para el usuario final. Los criterios de aptitud para los fines se pueden basar en algunas de las características descritas en el presente apéndice, pero en último término se expresarán como incertidumbre combinada aceptable.

A13. Variación de la matriz

En muchos sectores, la variación de la matriz es una de las fuentes de error en mediciones analíticas más importantes pero menos reconocida. Cuando definimos el sistema de análisis que se pretende validar, especificando, entre otras cosas, la matriz del material de ensayo, puede existir un considerable margen de variación de la matriz dentro de la categoría especificada. Por citar un ejemplo extremo, una muestra de la categoría de matriz “suelo” podría estar compuesta por arcilla, arena, caliza, laterita (principalmente Fe_2O_3 y Al_2O_3), turba, etc., o por una mezcla de estos materiales. Es fácil imaginar que cada uno de estos tipos de matrices podría producir un efecto específico en un método de análisis como la espectrometría de absorción atómica. Si no disponemos de información sobre el tipo de suelos que estamos analizando, los resultados estarán sometidos a una incertidumbre adicional debida a este efecto de variación de la matriz.

Las incertidumbres debidas a la variación de la matriz se deben cuantificar por separado, porque no se tienen en cuenta en otras partes del proceso de validación. La información se obtiene reuniendo un conjunto representativo de las matrices que probablemente encontremos en la categoría definida, en todos los casos con concentraciones de analito dentro del intervalo adecuado. Se analizan los materiales siguiendo el protocolo y se estima el sesgo de los resultados. La estimación del sesgo se deberá realizar habitualmente mediante adición y recuperación, excepto en el caso en que los materiales de ensayo sean MRC. La incertidumbre se estima mediante la desviación típica de los sesgos. (Nota: esta estimación comprenderá también una contribución a la varianza debida a la repetición de los análisis. Si se ha utilizado el método de adición, su magnitud será $2\sigma_r^2$. Si se necesita una estimación rigurosa de la incertidumbre, este término se debe deducir de la varianza debida a la variación de la matriz, para evitar que se contabilice por partida doble.)

A14. Incertidumbre de la medición

En el sistema formal de estimación de la incertidumbre de la medición, ésta se estima mediante una ecuación o modelo matemático. Los procedimientos descritos como validación de métodos tienen por objeto asegurar que la ecuación utilizada para estimar el resultado, en la que se tienen debidamente en cuenta errores aleatorios de todo tipo, es una expresión válida que comprende todos los efectos reconocidos y significativos que afectan al resultado. Por consiguiente, con una advertencia que se describe más adelante, la ecuación o “modelo” objeto de validación se puede utilizar directamente para estimar la incertidumbre de la medición. Esto se hace siguiendo principios establecidos basados en la “ley de propagación de la incertidumbre” que para efectos de insumos independientes es

$$u [y(x_1, x_2, \dots)] = \sqrt{\sum_{i=1, n} c_i^2 u(x_i)^2}$$

en que $y(x_1, x_2, \dots, x_n)$ es una función de varias variables independientes x_1, x_2, \dots , y c_i es un coeficiente de sensibilidad cuyo valor es $c_i = \partial y / \partial x_i$, la diferencial parcial de y con respecto a x_i ; $u(x_i)$ y $u(y)$ son *incertidumbres típicas*, es decir, incertidumbres de la medición expresadas en forma de desviaciones típicas. Dado que $u[y(x_1, x_2, \dots)]$ es una función de varias estimaciones de incertidumbres independientes, se conoce como *incertidumbre típica combinada*.

En consecuencia, para estimar la incertidumbre de la medición a partir de la ecuación $y = f(x_1, x_2, \dots)$ utilizada para calcular el resultado, es necesario determinar primero las incertidumbres $u(x_i)$ correspondientes a cada uno de los términos x_1, x_2 , etc.; combinar a continuación estas incertidumbres con los términos adicionales necesarios para representar los efectos aleatorios determinados en la validación y, finalmente, tener en cuenta los posibles efectos adicionales. El modelo estadístico implícito en la anterior argumentación sobre la precisión es

$$y = f(x_1, x_2, \dots) + \delta_{\text{run}} + e$$

en que e representa el error aleatorio de un resultado concreto. Como se sabe, por los experimentos de precisión, que los términos δ_{run} y e presentan respectivamente las desviaciones típicas σ_{run} y σ_r , estas últimas (o, de forma rigurosa, sus valores estimados s_{run} y s_r) son las incertidumbres asociadas a estos términos adicionales. En el caso en que se utilice el promedio de los distintos resultados dentro del proceso analítico, la incertidumbre combinada asociada a estos dos términos es (según la fórmula mencionada anteriormente) $s_{\text{tot}} = (s_r^2/n + s_{\text{run}}^2)^{1/2}$. Obsérvese que si se comprueba que los términos de precisión varían en función de la concentración de analito, el valor estimado de la incertidumbre de un determinado resultado debe utilizar el término de precisión adecuado para esa concentración. Por consiguiente, la estimación de la incertidumbre se basa directamente en el modelo estadístico adoptado y comprobado en la validación. A esta estimación se deben añadir los términos adicionales necesarios para tener en cuenta (en particular) la heterogeneidad y el efecto de la matriz (véase la sección A13). Por último, la incertidumbre típica calculada se multiplica por un “factor de cobertura”, k , obteniéndose una incertidumbre expandida, es decir: “un intervalo que previsiblemente abarca una gran porción de la distribución de valores atribuibles al mensurando” [8]. Cuando el modelo estadístico está sólidamente arraigado, y se sabe que la distribución es normal y el número de grados de libertad asociado al valor estimado es elevado, generalmente se selecciona un valor de k igual a 2. En este caso, la incertidumbre expandida corresponde aproximadamente a un intervalo de confianza del 95 %.

Se debe hacer una advertencia importante. Los ensayos realizados para comprobar la validez del modelo estadístico adoptado son necesariamente imperfectos. Se ha señalado ya que estos ensayos no pueden demostrar que un efecto sea exactamente igual a cero; sólo pueden demostrar que un efecto es demasiado pequeño para que pueda detectarse dentro del margen de incertidumbre asociado a la prueba de significación en cuestión. Un ejemplo particularmente importante es la prueba de significación del sesgo del laboratorio. Es evidente que si esta es la única prueba realizada para confirmar la veracidad, existirá alguna incertidumbre residual sobre si el método está verdaderamente sesgado o no. Por consiguiente, cuando estas incertidumbres son significativas con respecto a la incertidumbre calculada anteriormente, se debe prever un margen adicional.

En el caso de un valor de referencia incierto, el margen adicional más sencillo es la incertidumbre declarada del material, combinada con la incertidumbre estadística de la prueba aplicada. No se incluye en los objetivos del presente documento una explicación completa de la cuestión; la referencia 9 proporciona una descripción más detallada. No obstante, es importante señalar que si bien la incertidumbre estimada directamente del modelo estadístico adoptado es la incertidumbre *mínima* que puede asociarse a un resultado analítico, será con casi toda seguridad una subestimación de la misma; asimismo una incertidumbre expandida basada en las mismas consideraciones y con $k = 2$ no proporcionará una confianza suficiente.

La guía de la ISO [8] recomienda que, para mayor confianza, en lugar de añadir términos de forma arbitraria, se aumente el valor de k según sea necesario. La experiencia práctica sugiere que para estimaciones de

la incertidumbre basadas en un modelo estadístico validado, pero sin más evidencia que proporcione una confianza adicional en el modelo que la que proporcionan los estudios de validación, k no debe ser menor que 3. Cuando existan razones de peso para dudar de que el estudio de validación sea completo, k se debe aumentar ulteriormente, según sea necesario.

APÉNDICE B. CONSIDERACIONES ADICIONALES PARA LA ESTIMACIÓN DE LA INCERTIDUMBRE EN ESTUDIOS DE VALIDACIÓN

B1. Análisis de la sensibilidad

La expresión básica utilizada en la estimación de la incertidumbre

$$u [y(x_1, x_2, \dots)] = \sqrt{\sum_{i=1, n} c_i^2 u(x_i)^2}$$

requiere el uso de los “coeficientes de sensibilidad” c_i . En la estimación de la incertidumbre es frecuente comprobar que, si bien un determinado factor de influencia x_i tiene una incertidumbre conocida $u(x_i)$, el coeficiente c_i no está suficientemente caracterizado o no se puede obtener fácilmente a partir de la ecuación del resultado. Esto ocurre con particular frecuencia cuando un efecto no está incluido en la ecuación de la medición, porque normalmente no es significativo o porque la relación no se conoce suficientemente bien para justificar una corrección. Por ejemplo, pocas veces se caracteriza de forma detallada el efecto de la temperatura de la solución T_{sol} en un procedimiento de extracción a temperatura ambiente.

En el caso en que se desee evaluar la incertidumbre de un resultado asociada a este tipo de efecto, se puede determinar el coeficiente de forma experimental. Para ello, lo más sencillo es modificar x_i y observar el efecto sobre el resultado, según un método muy similar al aplicado en las pruebas básicas de robustez. En la mayoría de los casos, en un primer momento basta con seleccionar a lo sumo dos valores de x_i diferentes del valor nominal y calcular un gradiente aproximado a partir de los resultados observados. El gradiente proporciona un valor aproximado de c_i y se puede determinar a continuación el término $c_i \cdot u(x_i)$. (Obsérvese que este método permite demostrar de forma práctica si un posible efecto afecta o no de forma significativa a los resultados).

En este tipo de experimento, es importante que el cambio de resultado observado sea suficiente para obtener un cálculo fiable de c_i . Es difícil predecirlo con antelación. No obstante, dado un intervalo permitido para la magnitud influyente x_i , o una incertidumbre expandida para dicha magnitud, que se espera produzca un cambio insignificante, es evidentemente importante evaluar c_i a partir de un intervalo más amplio. Se recomienda, por consiguiente, que para una magnitud influyente con un intervalo esperado $\pm a$, (en que $\pm a$ puede ser, por ejemplo, el intervalo permitido, el intervalo de incertidumbre expandida o un intervalo de confianza del 95 %) en el experimento de sensibilidad se aplique, cuando sea posible, un cambio de al menos $4a$ para asegurar resultados fiables.

B2. Juicio profesional

No es raro observar que, si bien se reconoce un efecto y que el efecto puede ser significativo, no siempre es posible obtener un valor estimado fiable de la incertidumbre. En estas circunstancias, la guía de la ISO establece claramente que es preferible una estimación de la incertidumbre basada en el juicio de un profesional que no tener en cuenta la incertidumbre. De este modo, cuando no se dispone de una estimación de la incertidumbre de un efecto potencialmente importante, los analistas deben basarse en su propio juicio profesional para determinar una incertidumbre probable y aplicarla en la estimación de la incertidumbre combinada. La referencia 8 proporciona orientación adicional sobre la estimación de la incertidumbre basándose en el juicio profesional.

REFERENCIAS

1. W. Horwitz. "Protocol for the design, conduct and interpretation of method performance studies", *Pure Appl. Chem.* 60, 855, 864 (1988); revised version: W. Horwitz. *Pure Appl. Chem.* 67, 331–343 (1995).
2. M. Thompson and R. Wood. "The international harmonised protocol for the proficiency testing of (chemical) analytical laboratories", *Pure Appl. Chem.* 65, 2123–2144 (1993). (Also published in *J. AOAC Int.* 76, 926–940 (1993).
3. M. Thompson and R. Wood. "Harmonised guidelines for internal quality control in analytical chemistry laboratories", *Pure Appl. Chem.* 67 (4), 49–56 (1995).
4. M. Thompson, S. Ellison, A. Fajgelj, P. Willetts, R. Wood. "Harmonised guidelines for the use of recovery information in analytical measurement", *Pure Appl. Chem.* 71 (2), 337–348 (1999).
5. Council Directive 93/99/EEC on the Subject of Additional Measures Concerning the Official Control of Foodstuffs, O. J., L290 (1993).
6. Procedural Manual of the Codex Alimentarius Commission, 10th ed., FAO, Rome, 1997.
7. Precision of Test Methods, Geneva, ISO 5725 (1994). Previous editions were issued in 1981 and 1986.
8. Guide to the Expression of Uncertainty in Measurement, ISO, Geneva (1993).
9. A. Williams, S. L. R. Ellison, M. Roesslein (Eds.). *Quantifying Uncertainty in Analytical Measurement*, 2nd ed. (English) (2000), ISBN 0-948926-1595, available from LCC Limited, Teddington, London, or at Eurachem Secretariat, <http://www.eurachem.org/>.
10. International Vocabulary of Basic and General Terms in Metrology, ISO, Geneva (1993).
11. Validation of Chemical Analytical Methods, NMKL Secretariat, Finland, NMKL Procedure No. 4. (1996).
12. "EURACHEM Guide: The fitness for purpose of analytical methods. A Laboratory Guide to method validation and related topics", LGC, Teddington 1996. Also available from the EURACHEM Secretariat and Web site.
13. Statistics Manual of the AOAC, AOAC International, Gaithersburg, MD (1975).
14. An Interlaboratory Analytical Method Validation Short Course developed by the AOAC INTERNATIONAL, AOAC International, Gaithersburg, MD (1996).
15. "Text on validation of analytical procedures" International Conference on Harmonisation. Federal Register, Vol. 60, p. 11260, March 1 (1995).
16. "Validation of analytical procedures: Methodology", International Conference on Harmonisation, Federal Register, Vol. 62, No. 96, pp. 27463–27467, May 19 (1997).
17. "Validation of methods", Inspectorate for Health Protection, Rijswijk, The Netherlands, Report 95-001.
18. A Protocol for Analytical Quality Assurance in Public Analysts' Laboratories, Association of Public Analysts, 342 Coleford Road, Sheffield S9 5PH, UK (1986).
19. "Validation of analytical methods for food control", Report of a Joint FAO/IAEA Expert Consultation, December 1997, FAO Food and Nutrition Paper No. 68, FAO, Rome (1998).
20. "Estimation and expression of measurement uncertainty in chemical analysis", NMKL Secretariat, Finland, NMKL Procedure No. 5 (1997).
21. M. Thompson and P. J. Lowthian. *J. AOAC Int.* 80, 676–679 (1997).
22. "Nomenclature in evaluation of analytical methods, including quantification and detection capabilities", IUPAC Recommendation, *Pure Appl. Chem.* 67, 1699–1723 (1995).
23. ISO 11843. "Capability of detection", (several parts), International Standards Organisation, Geneva.

24. M. Thompson. *Analyst* 125, 2020–2025 (2000).
25. M. Thompson. “Recent trends in inter-laboratory precision at ppb and sub-ppb concentrations in relation to fitness for purpose criteria in proficiency testing”, *Analyst* 125, 385–386 (2000).
26. “How to combine proficiency test results with your own uncertainty estimate—the zeta score”, Analytical Methods Committee of the Royal Society of Chemistry, AMC Technical Briefs No. 2, M. Thompson (Ed.), www.rsc.org/lap/rsccom/amc